A Stalnakerian Analysis of Metafictive Statements

Semeijn, Merel

Published in:
Proceedings of the 21st Amsterdam Colloquium

Publication date:
2017

Proceedings of the 21st Amsterdam Colloquium

Cremers, A.M.E.; van Gessel, T.; Roelofsen, F.

Proceedings of the 21st Amsterdam Colloquium

21st Amsterdam Colloquium 2017

Edited by Alexandre Cremers,
Thom van Gessel & Floris Roelofsen

Foreword

This is a collection of papers presented at the 21^{st} Amsterdam Colloquium, organized by the Institute for Logic, Language, and Computation (ILLC) at the University of Amsterdam, December 20–22, 2017. The bi-annual Amsterdam Colloquia aim at bringing together linguists, philosophers, logicians, cognitive scientists and computer scientists who share an interest in the formal study of the semantics and pragmatics of natural and formal languages.

Besides the regular programme, the 2017 edition featured two workshops on *Causality and Semantics* and *Formal and Distributional Perspectives on Meaning*, respectively, and one evening lecture, jointly organized with the E.W. Beth Foundation. The programme included eight invited talks and 47 contributed talks.

We would like to thank the members of the programme committee and all the reviewers, listed below, for their efforts in selecting the contributed talks. We would also like to thank Patty den Enting, Luca Incurvati, and Peter van Ormondt for their help in organizing the colloquium.

Lastly, we would like to thank the ILLC, the E.W. Beth Foundation, the Netherlands Organization for Scientific Research (NWO), and the European Research Council (ERC) for financial support.

Alexandre Cremers
Thom van Gessel
Floris Roelofsen

Programme Committee

General programme

Robert van Rooij (chair)	ILLC, University of Amsterdam
Maria Aloni	ILLC, University of Amsterdam
Franz Berto	ILLC, University of Amsterdam
Paul Dekker	ILLC, University of Amsterdam

Workshop on Causality and Semantics

Franz Berto	ILLC, University of Amsterdam
Peter Hawke	ILLC, University of Amsterdam
Robert van Rooij	ILLC, University of Amsterdam
Katrin Schulz	ILLC, University of Amsterdam

Workshop on Formal and Distributional Perspectives on Meaning

Raquel Fernández Rovira	ILLC, University of Amsterdam

Reviewers

Luis Alonso-Ovalle	McGill University
Daniel Altshuler	Heinrich-Heine-Universitt Düsseldorf
Scott Anderbois	Brown
Nicholas Asher	IRIT, Université Paul Sabatier, Toulouse
Alan Bale	Concordia University
Moshe E. Bar-Lev	The Hebrew University of Jerusalem
Raffaella Bernardi	University of Trento
Bronwyn Bjorkman	Queen's University
Adrian Brasoveanu	UC Santa Cruz
Lisa Bylinina	Meertens Instituut
Ivano Caponigro	University of California San Diego
Lucas Champollion	New York University
Simon Charlow	Rutgers University
Emmanuel Chemla	LSCP, ENS, CNRS, Paris
Ivano Ciardelli	ILLC, University of Amsterdam
Ariel Cohen	Ben-Gurion University of the Negev
Cleo Condoravdi	Stanford University
Elizabeth Coppock	University of Gothenburg
Luka Crnič	The Hebrew University of Jerusalem
Chris Cummins	The University of Edinburgh
Kathryn Davidson	Harvard University
Christopher Davis	University of the Ryukyus
Henriette De Swart	Utrecht University
Marco Del Tredici	ILLC, University of Amsterdam
Jakub Dotlačil	University of Groningen

Regine Eckardt	University of Konstanz
Tim Fernando	Trinity College Dublin
Michael Franke	University of Tbingen
Alexander Goebel	University of Massachusetts at Amherst
Zsofia Gyarmathy	Heinrich Heine Universität
Valentine Hacquard	University of Maryland
Andreas Haida	The Hebrew University of Jerusalem
Robert Henderson	University of Arizona
Wesley Holliday	University of California, Berkeley
Julie Hunter	Universitat Pompeu Fabra, Barcelona and Université Paul Sabatier, Toulouse
Sabine Iatridou	Massachusetts Institute of Technology
Thomas Icard	Stanford University
Luca Incurvati	ILLC, University of Amsterdam
Gerhard Jaeger	University of Tuebingen
Stefan Kaufmann	Department of Linguistics, University of Connecticut
Magdalena Kaufmann	University of Connecticut
Chris Kennedy	University of Chicago
Hadas Kotek	New York University
Angelika Kratzer	University of Massachusetts at Amherst
Manuel Križ	Institut Jean Nicod, École Normale Supérieure
Jeremy Kuhn	Institut Jean Nicod, École Normale Supérieure
Fred Landman	Tel Aviv University
Daniel Lassiter	Stanford University
Sven Lauer	University of Konstanz
Emar Maier	University of Groningen
Edwin Mares	Victoria University of Wellington
Salvador Mascarenhas	École Normale Supérieure, Department of Cognitive Studies
Louise McNally	Universitat Pompeu Fabra
Sarah Murray	Cornell University
Reinhard Muskens	Tilburg Center for Logic and Philosophy of Science
Rick Nouwen	Utrecht Institute for Linguistics OTS
Edgar Onea	University of Graz
Paul Portner	Georgetown University
Jessica Rett	University of California, Los Angeles
Craige Roberts	The Ohio State University
Maribel Romero	University of Konstanz
Jacopo Romoli	university of ulster
Mats Rooth	Cornell University
Daniel Rothschild	Columbia University
Gillian Russell	The University of North Carolina at Chapel Hill
Kjell Johan Sæbø	University of Oslo
Julian Schlöder	ILLC, University of Amsterdam
Bernhard Schwarz	McGill University
Yael Sharvit	University of California, Los Angeles
Benjamin Spector	Institut Jean Nicod, École Normale Supérieure
William Starr	Cornell University
Shane Steinert-Threlkeld	ILLC, University of Amsterdam

Martin Stokhof	ILLC, University of Amsterdam
Yasutada Sudo	University College London
Eric Swanson	University of Michigan
Kristen Syrett	Rutgers University
Anna Szabolcsi	New York University
Jakub Szymanik	University of Amsterdam
Lyn Tieu	Macquarie University
Bob van Tiel	Bielefeld University
Kai von Fintel	Department of Linguistics & Philosophy, MIT
Klaus Von Heusinger	Universität zu Köln
Galit W. Sassoon	Bar Ilan University
Matthijs Westera	Universitat Pompeu Fabra, Barcelona
Aaron Steven White	University of Rochester
Yoad Winter	Utrecht University
Yimei Xiang	Harvard University
Seth Yalcin	Berkeley
Hedde Zeijlstra	University of Göttingen
Ede Zimmermann	Goethe Universität Frankfurt
Sarah Zobel	Eberhard-Karls Universität Tuebingen
Willem Zuidema	ILLC, University of Amsterdam

Contents

Invited

Workshop: Causality and Semantics

Workshop: Formal and Distributional Perspectives on Meaning

Contributed

A Trivalent Approach to Anaphora and Presupposition*

Daniel Rothschild

University College London
d.rothschild@ucl.ac.uk

Abstract

This paper presents an alternative to standard dynamic semantics. It uses the strong Kleene connectives to give a unified account of e-type anaphora and presupposition projection. The system is more conservative and simple than standard dynamic treatments of these two phenomena, and, I argue, has empirical advantages in its treatment of disjunction and negation.

1 Unified accounts of anaphora and presupposition

The goal of this paper is to present a simple and novel system for capturing core data about anaphora and presupposition projection. With respect to presupposition there is no novelty: I simply use a variant of the strong Kleene trivalent logic to treat presupposition projection.[1] What is new is that I add some apparatus from dynamic semantics to extend the trivalent system to also cover e-type anaphora.

Heim, in her dissertation [1982], gave two treatments of e-type (donkey) anaphora. One (chapter 2) treated anaphora by means of explicit existential quantifiers in a fully static (and very standard) semantic framework, the other (chapter 3) introduced the first compositional dynamic semantics for anaphora. One of Heim's main arguments for adopting the second approach was that her dynamic system provided a unified treatment of anaphora and presupposition, something no other account provided. Heim's account of anaphora and presupposition has been modified and extended by, among others, Beaver [2001] into a unified and powerful system for the treatment of both.[2]

The dynamic treatment of presupposition projection has been criticized by Schlenker [2008, 2009] for its lack of explanatoriness. However, alternative treatments of presupposition projection such as Schlenker's local context approach and the trivalent approach do not obviously integrate well with an account of e-type anaphora.[3] In later work, Heim [1990] suggests integrating a static (presumably trivalent) presuppositional approach to definites with situation semantics and an e-type treatment of pronouns as disguised Fregean descriptions to cover donkey anaphora, a treatment elaborated in Elbourne [2005]. This paper is not the occasion for a full discussion of these semantic theories, but I will pause to note the following:

*I am indebted to Matt Mandelkern for extensive discussion.

[1] It is my view that, when the dust has settles, this remains the simplest viable treatment of presupposition projection on the market. See Peters [1979], Krahmer [1998], George [2007], Fox [2008, 2012] among others.

[2] A different tradition stemming from van der Sandt [1992] uses and Kamp's DRT to unify anaphora and presupposition. Beaver, to my mind, makes convincing arguments against this approach.

[3] I use *e-type anaphora* as a term to describe the general phenomenon in which pronouns are used without c-commanded antecedents, the relation between pronoun and antecedent being inter-sentential, across conditionals, or between the restrictor and matrix of an NP. An *e-type treatment*, by contrast, is a semantic account of such pronouns which treats them as akin to defined descriptions that have Russellian/Fregean semantics [such as, Cooper, 1979, Evans, 1977].

- Unlike standard dynamic accounts, these proposals have rarely, if ever, been spelled out in large fragments containing sentential connectives and negation.[4]
- The proposals contain complex definitions of quantifiers such as ⌜every⌝ with multiple layers of existential and universal quantification over individuals and situations (a property shared by most dynamic approaches but not by the account I present here).
- The connections between e-type anaphora and presupposition projection are rarely made explicit in this tradition.

For these and other reasons I do not see the situation-theoretic e-type approach as a particularly promising line for an integrated account of e-type anaphora and presupposition.

As I see the current situation, then, dynamic approaches provide the best unified accounts of presupposition and anaphora. So why should we bother rethinking the framework of dynamic semantics when it is so successful in this respect? Shouldn't we just accept its successes and move on, either just replacing it or expanding it, rather than tweaking? Here I stand with Dekker and Schlenker, in particular, who have suggested that the successes of dynamic semantics may not adequately motivate its foundational ideas.

For instance, a salient feature of standard dynamics semantics is to treat the semantic values of sentences not as truth-conditions but rather as context change potentials (CCPS).[5] In other words, instead of having semantic values be functions from points of a context to truth-values, semantic values are functions from contexts (sets of points) to contexts.[6] There are many obviously inexpressible such functions: for example, we do not have a sentences in any language that expresses the context change that moves any context to one which only accepts the fact that there are pink elephants. There are no knock-down considerations in favor of having lower-type semantic values, but lower types are simpler and, thus, all else equal to be preferred.

All else is never equal, though, and type-theoretical considerations are not my only ones. Another way in which my semantics is simpler is that the definitions of the quantifiers and connectives I use are essentially their classical definitions: the dynamic effects of these really do follow from their classical definitions (and the strong Kleene logic). Thus, I share the motivations for Schlenker's non-dynamic account of presupposition projection which relies on a classical understanding of connectives and quantifiers. This, again, adds to the simplicity of the semantic system and relatedly its learnability. More significant, perhaps, are empirical advantages: I handle the behavior of anaphora under double negation and through disjunctions in a straightforward way, something dynamic accounts tend to struggle with.[7]

My account is in the spirit of the constructive criticisms of dynamic semantics put forward by Dekker [1994, 2012] and Schlenker [2008, 2009]. The account is similar to Dekker's Predicate Logic with Anaphora (and is directly inspired by it), in that it also uses many of the conceptual innovations of dynamic semantics without resorting to a context change potential-based semantics. On the other hand, the account is parallel to Schlenker's static accounts of presupposition (transparency theory and his local context theory) in that it uses more standard, non-stipulative

[4] For example the fragment in Elbourne [2005]—often pointed to as one of the most extensive situation-theoretic-cum-descriptive treatments of donkey anaphora—does not contain a semantics for negation, and it is non-trivial to see how one can be added.

[5] Or in extensional fragments such as Groenendijk and Stokhof [1991] as functions from assignment functions to assignment functions rather than assignment functions to truth values.

[6] I am assuming here that contexts are sets whose elements I call points, these points can be worlds as in Stalnaker's framework or world-assignment function pairs as in many dynamic accounts.

[7] In some aspects of this, I follow Krahmer [1998], except that Krahmer combines DRT with trivalence, rather than simply having a trivalent system and he does not cover all the aspects of disjunction that I do.

definitions of all quantifiers and connectives, including conjunction. From a broader perspective, while Dekker treats e-type anaphora but not presuppositon, and Schlenker treats presupposition but not e-type anaphora, I try to treat both.

2 Some rules of the game

We will adopt a Heimian notion of context according to which a context is a set of pairs of assignment functions and worlds.[8] This conception, of course, does not commit us to higher-type semantic values, just the (fairly) uncontroversial idea that speaker and hearers keep track of possible discourse references for certain 'variables' introduced in discourse.[9]

Truth conditions—certain forms of irrelevant context-sensitivity aside—are simply functions from elements of such contexts (pairs of assignment functions and worlds) to truth values. Our semantics will be static or truth-conditional in that the semantic value of sentences will be such truth-conditions.[10]

The update rule associated with a sentence ϕ will be Stalnakerian [Stalnaker, 1970, Rothschild and Yalcin, 2015]. When a sentence ϕ is asserted in a context c we remove from c every element on which ϕ is not true.[11]

3 E-type pronouns and presuppositions

The following examples illustrate a small part of the connection between presupposition projection and e-type anaphoric relations.

(1) a. John used to smoke and he hasn't stopped smoking.
b. ?John hasn't stopped smoking and he used to smoke.
c. ?John didn't used to smoke and he hasn't stopped smoking.

(2) a. A man walked in and he wasn't wearing a hat.
b. ?He wasn't wearing a hat and a man walked in.
c. ?A man didn't walk in and he was wearing a hat.

It is worth sketching an aspect of the empirical connection between presupposition and anaphora. Consider a case where we have a complex sentence s with constituents ϕ and ψ such that ϕ presupposes X and ψ classically entails X. For example in (1-a), $\phi = \ulcorner$he hasn't stopped smoking$\urcorner$, $\psi = \ulcorner$he used to smoke$\urcorner$ and $X = \ulcorner$he used to smoke.$\urcorner$ If s does not itself

[8] This is also the approach explored *inter alia* in Groenendijk et al. [1996]: context includes information about variable assignments, not just worlds

[9] This is in the spirit of Lewis [1983] and even seems to be countenanced in an unsystematic pragmatic way by Stalnaker [1998]. Note that this notion can lead to a flavor of discourse dynamism within a completely static semantic system as discussed by Rothschild and Yalcin [2016, §5].

[10] It is important to distinguish this kind of static compositional semantics from what Rothschild and Yalcin [2015] call dynamics at the *conversational level*. That is, the CCPs associated with assertion in our semantic/pragmatic whole may be dynamic in various senses, but the sentences themselves are static in that they are not functions from contexts to contexts but rather simply truth assignments for individual points.

[11] Of course, the effect of this update rule will not always be what Stalnaker had in mind in the classic papers where he suggests these updates since the context includes not just worlds but also assignment functions. I share with the dynamic tradition a lack of interest in the question of whether sentences express propositions. Since contexts are not just sets of worlds we cannot simply identify the semantic value of a sentence with the set of worlds it is true in. But for reasons addressed already in Lewis [1980] that is not generally viable in semantic theorizing. Certainly we can define various different notions of content in contexts, but we need not take a stand on these.

presuppose X, and this is because of ψ, then we'll say that ψ allows for local satisfaction of the presupposition of ϕ (it filters it out). This is the case in (1-a). Very roughly speaking, the same configurations that allow for local satisfaction of presuppositions also allow for e-type anaphora. This is what is illustrated by the examples above.

Given this connection between local satisfaction of presupposition and e-type anaphora we might expect a theoretical connection. Our account, like the dynamic account stemming from Heim [1982], tries to make good on that expectation.

4 A trivalent account of presupposition

Let me give a brief outline of how the facts about presupposition projection can be accounted for on a trivalent framework. On a trivalent semantics, sentences can be either true, false, or undefined (1,0, or #). The connectives handle undefinedness according to the strong Kleene truth tables, the guiding principle of which is to give truth values when those are determined by what is defined. We can also add order effects, more closely matching standard theories of presupposition projection by using Peter's truth tables [Peters, 1979, Krahmer, 1998, George, 2007].

The relationship between the trivalent truth tables and context is important here. A sentence ϕ is acceptable in a context c iff ϕ is true or false at every element of c. This, sometimes called Stalnaker's principle, was proposed by him as an intuitive principle, but has since been recognized to be rather a kind of stipulation [for discussion, see Soames, 1989, Rothschild, 2008b, Fox, 2012].

To give an example, let us treat ⌜John stopped smoking⌝ as undefined if John didn't use to smoke and true or false otherwise. Then, given the strong Kleene understanding of conjunction, (1-a) will be defined in any context. This is because when the presupposition of the second conjunct is not satisfied the first conjunct is false and so the entire sentence is undefined.

5 E-type Anaphora and Content

What has been called the problem of the formal link [Heim, 1990] puts a particularly sharp constraint on how we treat anaphoric connections. Here are types of examples due to Partee and Heim respectively:

(3) a. ? Nine of the ten marbles are on the floor. . . . ? It's on the couch.
b. One of the ten marbles is not on the table. . . . It's on the couch.

(4) a. ? Every married man loves her.
b. Every man with a wife loves her.

The lesson here is that pronouns without appropriately marked NP antecedents are difficult, and often result in infelicity. This suggests that if we are to use a presuppositional approach we cannot rely on simply presuppositions about the state of the world. Rather we also need presuppositions that somehow involve variables.[12]

[12]The modern e-type treatment of anaphora rather attempts to cover such facts by positing syntactic conditions on the licensing of covert descriptions, such as Elbourne's [2005] NP-deletion. My view is that such conditions face serious challenges as they separate the syntactic licensing conditions from the semantics of the pronouns and will inevitable make bad predictions. Here is one example:

(i) ?Everyone who doesn't have a home but knows a home-owner, stays in it.

6 Anaphoric presuppositions: intersentential case

Let's focus for a moment on the *inter*-sentential case, such as (3-a) and (3-b). The question is how the context differs between these two examples to make the assertions different: that is, what effect does the first sentence have on the context to make the pronoun in the second sentence felicitous. Let us assume, as in the Heim's approach, that pronouns are simply variables. What we need to explain the contrasts in the section above is to posit special constraints on the use of variables. For example, Heim puts a *familiarity presupposition* on the use of definite variables—a special structural condition on the local context of a pronoun. In this section I outline a related approach that fits more naturally with a standard trivalent treatment of presuppositions.

I will assume something like the partiality of assignment functions, though in a slightly non-standard way. What I will assume is that assignment functions are functions from variables onto the usual domain D as well as an absurd object $\perp$. Empty contexts include all assignments.[13]

The presupposition of the use of a definite variable x is simply that x does not refer to $\perp$. To get this to work we need the harmless assumption that the extension of every predicate is undefined when applied to $\perp$.

What about our treatment of indefinites? How do they ensure that intersentential anaphora works? Nothing special is needed, except assuming, as Heim does, that indefinites put constraints on variables rather than having existential force in the usual sense. ⌜One marble$_x$ is not on the table⌝ is true at $\langle f, w\rangle$ iff $f(x)$ is a man who walked in at w. Any context on which it is true at every point, thus, will make ⌜it$_x$'s on the couch⌝ defined.

Let's go through a simple example with truth conditions spelled out:

$$[\![\text{A man}_x \text{ walked in.}]\!]^{f,w} = \begin{cases} 1 \text{ iff } f(x) \text{ is a man who walked in in } w \\ 0 \text{ otherwise} \end{cases}$$

$$[\![\text{He}_x \text{ had a drink}]\!]^{f,w} = \begin{cases} \# \text{ if } f(x) = \perp \\ 1 \text{ if } f(x) \text{ had a drink in } w \\ 0 \text{ otherwise} \end{cases}$$

The point is that once the context absorbs the first sentence, then the second sentence is guaranteed to be acceptable since all the points at which $f(x) = \perp$ have been eliminated. In terms of truth-conditions, this theory so far matches Heim: indefinites have existential force given that the empty assignment allow all possible assignments.[14]

7 Conjunction

The story for inter-sentential anaphora extends to a treatment of e-type anaphora across a conjunction:

(5) A man$_x$ walked in and he$_x$ ordered a drink.

(ii) Every who isn't a home-owner but knows someone who has a home, stays in it.

The problem is that it would seem that ⌜a home⌝ in (i) should license the description ⌜the home⌝ whose presupposition is then satisfied by restrictor state.

[13] On the standard use of partial assignment functions in dynamic semantics the empty assignment is that in which no variable has a defined assignment, not one in which every possible assignment (including empty/absurd ones) is in the context.

[14] While we have not yet put any condition on the use of indefinites, we might assume that indefinite should not be reused because of some variation of a maximize presupposition rule [Heim, 1991].

At any point in the context in which x is assigned $\perp$ the sentence is false on the strong Kleene understanding of conjunction since the first conjunct is false.[15] If we wish to explain why the reverse order, as in (6), is infelicitous will need to use the Peters [1979] version of the connectives or apply an order constraint.

(6) ?He$_x$ ordered a drink and a man$_x$ walked in.

8 Taking stock

Let us take stock of where we are so far. Here are the salient features of the semantics and pragmatic that have posited so far: a) Heim's basic understanding of context as sets of pairs of worlds and assignment functions, b) a particular type of assignment functions that includes absurd objects, c) a variable-based semantics for both indefinites and pronouns, d) a trivalent logic and strong Kleene connectives, e) Stalnaker's updates rule (i.e. pointwise) and principle for presupposition felicity (i.e. a sentence is felicitous if it is defined at every point in the context). With these resources we give a reasonable treatment of presupposition, intersentential anaphora as well as anaphora across conjunctions. These are the easy cases, however: the challenge will be to give adequate treatments of negation, disjunctions, quantifiers, and adverbs of quantification.

(One thing to note is that this semantics does not behave well when the same index is reused by a new quantifier. Different dynamic systems have treated reused indices in different ways. My attitude is that as there is no empirical evidence that indices are ever reused in natural language not choose systems on the basis of how they respond to reused indices.)[16]

9 Negation?

Let us start with negation. The most obvious problem is with the negation of indefinites. Recall that we wanted $\ulcorner$A man$_x$ walks in$\urcorner$ to be true at a point iff x is a man who walks in at that point. This is necessary in order that the sentence does the job of satisfying later presuppositions of variables: the context, once the sentence has been asserted, needs to be one that includes the fact that x picks out a man who walked in. What about the (wide-scope) negation of $\ulcorner$A man$_x$ walks in$\urcorner$? In an empty context this eliminates any world in which any man walks in.

If we're going to treat truth-value gaps as presupposition-invoking we also need $\ulcorner$A man$_x$ walks in$\urcorner$ to be true or false everywhere. So, if we stick with a trivalent logic with a strong Kleene negation we are in a bind, since our truth-conditions have in fact forced our hands with our falsity-conditions, and these are not what we want. It is exactly these kinds of considerations which led Heim in her static fragment of chapter 2 to propose an existential closure operation under negation.

We cannot, of course, simply existentially close all variables, since pronouns do not undergo existential closure under negation. One option, which we will take here, is to simply existentially close those variables that are not at risk of causing presupposition failures. There are some subtleties here, but we will use the following definitions which, with some relatively harmless auxiliary assumptions, should prove adequate. We will say a sentence ϕ is *definedness-sensitive* to a variable x iff there exists a world w and an assignment function f s.t.$[\![\phi]\!]^{f_{x\to\perp},w} = \#$ and

[15] See the appendix for details including the strong Kleene conjunction.

[16] Here I'm in agreement with an unpublished paper by Charlow [2016]. Note also that we can probably explain why do not generally coindex two indefinite quantifiers by means of a maximize presupposition rule.

for all o, $[\![\phi]\!]^{f_{x\to o},w} \neq \#$.[17] We say that $f'[\phi]f$ iff f' agrees with f on all definedness-sensitive variables.

We can now define our dynamic existential closure operator, † as follows:

$$[\![\dagger\phi]\!]^{f,w} = \begin{cases} 1 \text{ if there is an } f'[\phi]f \text{ s.t. } [\![\phi]\!]^{f',w} = 1 \\ 0 \text{ if for all } f'[\phi]f,\ [\![\phi]\!]^{f',w} = 0 \\ \# \text{ otherwise} \end{cases}$$

Now, if we have a strong Kleene negation we can treat the negated sentences as follows:

$$[\![\neg\,\dagger \text{ A man}_x \text{ walked in}]\!]^{f,w} = \begin{cases} 1 \text{ iff no man in } w \text{ walked in} \\ 0 \text{ otherwise} \end{cases}$$

The closure or assertion operator † will prove very useful in many places (including with disjunctions and under quantifiers), so we need not think of it as merely required for negation. For now, we will assume it can be freely placed under other operators (at the root level it only eliminates anaphoric potentials so is not useful).

Moreover, we also now can get anaphora across double negations by not using † at all, e.g. $[\![\neg\neg\text{A man}_x \text{ walked in}]\!] = [\![\text{A man}_x \text{ walked in}]\!] \neq [\![\neg\neg\,\dagger \text{ A man}_x \text{ walked in}]\!]$

What about $[\![\neg\,\dagger\,\neg\text{A man}_x \text{ walked in}]\!]$? This has the following truth-conditions:

$$= \begin{cases} 1 \text{ if some man in } w \text{ walked in} \\ 0 \text{ otherwise} \end{cases}$$

It simply lacks anaphoric potential.

What about negation without the †-operator: $\ulcorner\neg$A man$_x$ walked in$\urcorner$? While this does successfully put conditions on x, it does not ensure that $x \neq \perp$. In addition, when asserted in a context without any variable information it does not put any worldly conditions on the context. We might hope to eliminate such parses on pragmatic grounds, or postulate syntactic constraints to remove them.

10 Disjunction

10.1 Partee disjunction

We have already seen that our theory accounts naturally for anaphoric connections across double negations. What about the related question of how the theory accounts for this kind of disjunction example, due to Partee:

(7) There isn't a bathroom here, or it's under the stairs.

We don't naturally get a coherent reading. For on the parse in (8) the entire sentence will presuppose that x is assigned.

(8) $(\neg\,\dagger$ there is a bathroom$_x$ $) \vee$ (it$_x$'s under the stairs)

It is easy to check that there is no way to place the † operator to yield the desired reading.

However, if we are allowed to insert logically redundant material (in a classical sense), then we *can* get the desired reading. Note that from the perspective of propositional logic $\ulcorner\phi \vee \psi\urcorner$ is equivalent to $\ulcorner\phi \vee (\neg\phi \wedge \psi)\urcorner$. So from a classical perspective $\ulcorner(\neg$there is a bathroom$_x) \vee$ (it$_x$'s under the stairs)$\urcorner$ is equivalent to $\ulcorner(\neg$there is a bathroom$_x$ $) \vee$ (there is a bathroom$_x \wedge$ it$_x$'s under the stairs)$\urcorner$. Now if we just add a † operator we get the correct reading: $\ulcorner(\neg\,\dagger$ there is a bathroom$_x$ $) \vee$ (there is a bathroom$_x \wedge$ it$_x$'s under the stairs)$\urcorner$.

[17] A problem is this: $\ulcorner\exists x(x = x \vee$ John knows $1{+}1 = 2)\urcorner$.

What I am proposing is that we can tweak logical forms not only by adding the †-operator, but also by adding (classically) logically redundant conjunctions. The combination of these two free operations will then give us the desired readings under disjunctions. Of course, such free operations provide a significant divergence between overt syntactic form and that which finally makes up the meaning, but the proposed operations are sufficiently constrained, I believe, to be plausible.[18]

10.2 Stone disjunction

Another aspect of the dynamics of disjunction can be handled in my system without modification. Consider disjunctions that can serve as anaphoric antecedents to donkey anaphora:

(9) Either a man will bring a comb or a woman will bring a brush. In either case, ask them to leave it for me.

One natural suggestion is that the pronouns are linked to both antecedents as follows:[19]

(10) Either a man$_x$ will bring a comb$_y$ or a woman$_x$ will bring a brush$_y$. In either case, ask them$_x$ to leave it$_y$ for me.

Stone [1992] posed examples of this general form as a particular problem for dynamic semantics. What we see, though, is that in our semantics with a simple classical semantics for disjunction we have no problem with these examples. For any context in which the first sentence is accepted both x and y will not refer to $\perp$ but rather to either a man or woman or a comb or brush, respectively. Thus, the presuppositions of the pronouns in the second sentence will be satisfied, yielding the desired interpretation. Again, we see that by using a trivalent semantics without stipulative accessibility rules we eliminate problems that plague traditional dynamics semantics.

10.3 From disjunction to conditional

Consider this kind of anaphoric connection:

(11) Either it's a holiday or a customer$_x$ will come in. And if it's not a holiday, he$_x$'ll want to be served.

To my knowledge, cases such as (11) have not been discussed in the literature, though they resemble, in some respects, cases of modal subordination. Standard dynamic accounts have no natural resources to account for them, while e-type approaches can easily treat them since the presuppositions of a definite description such as ⌜the customer⌝ is satisfied in the local context of the consequent of the conditional. Likewise the account I am advocating here naturally captures such examples since the special presupposition of the pronoun (that x does not pick out $\perp$) is conditionally satisfied once the context is updated with the first disjunct.[20]

[18] Kamp and Reyle [1993] also consider adding extra material to the second disjunct to get the desired reading, but they do not give an explicit system. I am recycling the basic idea from [Rothschild, 2008a] of facilitating dynamic effects by allowing reconstructions of logical form according to classical equivalence. While I describe such operations as free here, I believe we will need some constraints on them in order not to generate unattested readings. The viability of this proposal will ultimately depend on the nature of these constraints.

[19] Schlenker [2011] gives evidence from sign language that this is the logical form of anaphora with disjunctive antecedents.

[20] I make the simplifying assumption here that the conditional in the second sentence is just a material conditional with strong Kleene semantics.

11 Quantifiers

With respect to generalized quantifiers such as ⌜every⌝ we will assume a classical behavior. The syntactic/semantic assumptions of our quantifiers are relatively simple: a quantifier Q_x takes two arguments of sentential type, which are assumed to contain the free variable x. Q_x then expresses a conservative relationship between the objects satisfying the two arguments (i.e. the objects that when the assignment function assigns x to those objects makes the arguments true), as in standard generalized quantifier theory [Barwise and Cooper, 1981].

The critical problem we face is how to understand anaphoric relationships between the restrictor and matrix of quantifiers. Consider donkey sentences:

(12) Every (man who owns a donkey, beats it)

Given our reformation rules across logical equivalence, these sentences present no problem. For conservativity ensures that conjoining the restrictor to the matrix in the standard fragment makes no difference to truth conditions. So we switch the logical form of (12) to (13).

(13) Every († man who owns a donkey, † (man who owns a donkey and beats it))

This gets us what is called the 'weak' reading of donkey anaphora, namely that every man who owns a donkey beats at least one donkey he owns. We will need to avail ourselves of one of the many strategies in the literature for also obtaining the other reading, but I leave that for another occasion.[21]

It is notable that the same technique, using classically equivalent logical forms to capture anaphoric relations, works for both Partee-disjunction and classic donkey anaphora under quantifiers.

12 Dynamic adverbs of quantification

As it happens I think the correct treatment of adverbs of quantification requires situational quantifiers, for roughly the reasons discussed in von Fintel [1994]. However, if we want to define a Lewisian adverb of quantifier that behaves appropriate for examples like these there is no technical obstacle. I give a definition in the appendix which follows the usual dynamic definitions of adverbs of quantifiers as Lewisian [1975] unselective quantifiers.

13 Summary and comparative remarks

Our proposed semantics took a number of important ideas from the literature on dynamic semantics, particularly Heim's dissertation.

- Contexts have a Heimian file structure: they are sets of assignment function world/pairs.
- Pronouns and indefinites put conditions on variables.
- There is kind of default existential quantification at the sentence level and under operators such as negation.

[21] In my view the weak reading is the right one to get as it is always attested whereas some sentences with donkey anaphora have no ambiguity:

(i) No man who owns a donkey beats it.

It is worth comparing this to a dynamic semantics that also covers anaphora and presupposition. As I noted earlier Beaver's [2001] ABLE is the obvious comparison point as it covers presupposition and anaphora (as well as epistemic modals) and builds on much other work in dynamic semantics from Heim [1982], Kamp [1981] onwards. There are a number of significant differences between my system and Beaver's (and other dynamic systems):

My system has semantic values that are functions from assignment world pairs to truth values. Beaver's are functions from sets of assignment worlds pairs to sets of assignment worlds pairs (or relations in type, but he only uses functional relations in his fragment). Beaver's definitions of connectives and quantifiers all make reference explicitly to order effects of accessibility relations. My quantifiers and connectives are simply those from a strong Kleene logic.

On the other hand, my system allows free insertion of †-operators, giving existential closures as well restricted additions of conjunctives where they do not (classically) affect the truth-conditions.

My system, in terms of type, is close to that of Dekker [1994, 2012] who also assigns truth-conditions as semantic values rather than CCPs. However, unlike Dekker, I provide a treatment of presupposition and have a more classical treatment of quantification and connectives. In my rigid use of standard (trivalent) quantifier definitions this proposal is in the spirit of Schlenker's work on presupposition.

A Syntax

The sets V of variables: $x, y, z \ldots$
Relational predicates, $P, R, Q \ldots$
Where P is a relational predicate and $\gamma_1 \ldots \gamma_n$ are variables, ϕ and ψ are arbitrary wff, we form wff as follows:
$P(\gamma_1 \ldots \gamma_n)|\text{some}_{\gamma_1}(\phi,\psi)|\text{every}_{\gamma_1}(\phi,\psi)|\phi \wedge \psi|\phi \vee \psi|\neg\phi|\text{always}(\phi,\phi)|\dagger\,\phi$

B Semantics

Let D be the domain of objects and W be a set of worlds. Let an assignment function be a function from V to $D \cup \perp$, where $\perp$ is a special object not in the domain. An interpretation I is a mapping from relational predicates and worlds to n-tuples of D. The denotation function $[\![\cdot]\!]$ is a function from a wff, an interpretation, an assignment function and world to the set $\{0, 1, \#\}$. (We generally do not refer to the interpretation function, but just refer to predicates holding in worlds as usual.)

We let $f[\phi]f'$ iff f agrees with f' on all variables x such that there exists a world w and an assignment function g s.t. $[\![\phi]\!]^{g_{x\to\perp},w} = \#$ and for all o, $[\![\phi]\!]^{g_{x\to o},w} \neq \#$

$$[\![P(\gamma_1,\ldots\gamma_n)]\!]^{f,w} = \begin{cases} \# \text{ if any of } f(\gamma_1)\ldots f(\gamma_n) = \perp \\ 1 \text{ iff } \langle f(\gamma_1),\ldots f(\gamma_n)\rangle \text{ is in the extension of } \phi \text{ at } w \\ 0 \text{ otherwise} \end{cases}$$

$$[\![\phi \wedge \psi]\!]^{f,w} = \begin{cases} 1 \text{ if } [\![\phi]\!]^{f,w} = 1 \text{ and } [\![\psi]\!]^{f,w} = 1 \\ 0 \text{ if } [\![\phi]\!]^{f,w} = 0 \text{ or } [\![\psi]\!]^{f,w} = 0 \\ \# \text{ otherwise} \end{cases}$$

$$[\![\phi \vee \psi]\!]^{f,w} = \begin{cases} 1 \text{ if } [\![\phi]\!]^{f,w} = 1 \text{ or } [\![\psi]\!]^{f,w} = 1 \\ 0 \text{ if } [\![\phi]\!]^{f,w} = 0 \text{ and } [\![\psi]\!]^{f,w} = 0 \\ \# \text{ otherwise} \end{cases}$$

$$[\![\neg\phi]\!]^{f,w} = \begin{cases} 1 \text{ if } [\![\phi]\!]^{f,w} = 0 \\ 0 \text{ if } [\![\phi]\!]^{f,w} = 1 \\ \# \text{ otherwise} \end{cases}$$

$$[\![\text{some}_x(\phi,\psi)]\!]^{f,w} = \begin{cases} 1 \text{ if } [\![\phi]\!]^{f,w} = 1 \text{ and } [\![\psi]\!]^{f,w} = 1 \\ 0 \text{ if } f(x) = \perp \text{ or } [\![\phi]\!]^{f,w} = 0 \text{ or } [\![\psi]\!]^{f,w} = 0 \\ \# \text{ otherwise} \end{cases}$$

$$[\![\text{every}_x(\phi,\psi)]\!]^{f,w} = \begin{cases} 1 \text{ if } \forall o \in D : [\![\phi]\!]^{f_{x\to o},w} = 1 \text{ and } [\![\psi]\!]^{f_{x\to o},w} = 1 \\ 0 \text{ if } \exists o \in D : [\![\phi]\!]^{f_{x\to o},w} = 1 \text{ and } [\![\psi]\!]^{f_{x\to o},w} = 0 \\ \# \text{ otherwise} \end{cases}$$

$$[\![\text{always}(\phi,\psi)]\!]^{f,w} = \begin{cases} 1 \text{ if } \forall f'[\phi]f \text{ such that } [\![\phi]\!]^{f',w} = 1, \exists f''[\psi]f' \text{ such that } [\![\psi]\!]^{f'',w} = 1 \\ 0 \text{ if } \exists f'[\phi]f \ [\![\phi]\!]^{f',w} = 1 \text{ and } \forall f''[\psi]f' \ [\![\psi]\!]^{f'',w} = 0 \\ \# \text{ otherwise} \end{cases}$$

$$[\![\dagger\phi]\!]^{f,w} = \begin{cases} 1 \text{ if } \exists f'[\phi]f \text{ such that } [\![\phi]\!]^{f',w} = 1 \\ 0 \text{ if } \forall f'[\phi]f \ [\![\phi]\!]^{f',w} = 0 \\ \# \text{ otherwise} \end{cases}$$

C Transformation mechanisms

In moving from expressed logical forms to the form interpreted we allow the following two alterations:

†**-insertion:** Replace any instance of a wff ϕ inside a wff with $\dagger\phi$

Adding redundant conjunctions: if a wff contains the wffs ϕ and ψ replace any instance of ψ with $\phi \wedge \psi$ if the replacement is a classically equivalent.(Definition of classical equivalence: formulas α and β are classically equivalent if when all † operators are removed for every interpretation I and assignment function f $[\![\phi]\!]^{f,w} = 1$ iff $[\![\psi]\!]^{f,w} = 1$.)

References

John Barwise and Robin Cooper. Generalized quantifiers and natural language. *Linguistics and Philosophy*, pages 159–219, 1981.

David Beaver. *Presupposition and Assertion in Dynamic Semantics*. CSLI, 2001.

Simon Charlow. Where is the destructive update problem? unpublished manuscript, Rutgers, 2016.

Robin Cooper. The interpretation of pronouns. In Frank Heny and Helmut S. Schnelle, editors, *Syntax and Semantics*, volume 10, pages 61–92. Academic Press, 1979.

Paul Dekker. Predicate logic with anaphora (seven inch version). In Lynn Santelmann and Mandy Harvey, editors, *Proceedings of SALT IV*, pages 79–95. Ohio State University, 1994.

Paul Dekker. *Dynamic Semantics*. Springer, 2012.

Paul Elbourne. *Situations and Individuals*. MIT Press, 2005.

Gareth Evans. Pronouns quantifiers and relative clauses (i). *Canadian Journal Of Philosophy*, 7(3):467–536, 1977.

Danny Fox. Two short notes on Schlenker's theory of presupposition projection. *Theoretical Linguistics*, 34:237–252, 2008.

Danny Fox. Presupposition projection from quantificational sentences: trivalence, local accommodation, and presupposition strengthening. manuscript, HUJI and MIT, 2012.

Benjamin George. Predicting presupposition projection: some alternatives in the strong Kleene tradition. manuscript, UCLA, 2007.

J. Groenendijk and M. Stokhof. Dynamic predicate logic. *Linguistics and Philosophy*, 14: 39–100, 1991.

Jeroen Groenendijk, Martin Stokhof, and Frank Veltman. Corefrence and modality. In Shalom Lappin, editor, *Handbook of Contemporary Semantic Theory*. Blackwell, 1996.

Irene Heim. *The Semantics of Definite and Indefinite Noun Phrases*. PhD thesis, University of Massachusetts, Amherst, 1982.

Irene Heim. E-type pronouns and donkey anaphora. *Linguistics and Philosophy*, 13:137–177, 1990.

Irene Heim. Artikel und definitheit. In A v. Stechow and D. Wunderlich, editors, *Semantics: An International Handbook of Contemporary Research*. de Gruyter, 1991.

Hans Kamp. A theory of truth and semantic representation. In J. Groenendijk, T.M.V. Janssen, and M.B.J. Stokhof, editors, *Formal Methods in the Study of Language*, pages 277–322. Mathematish Centrum, 1981.

Hans Kamp and Uwe Reyle. *From Discourse to Logic*. Kluwer, 1993.

Emiel Krahmer. *Presupposition and Anaphora*. CSLI, 1998.

David Lewis. Adverbs of quantification. In Edward L. Keenan, editor, *Formal Semantics of Natural Language*. Cambridge University Press, 1975.

David Lewis. Index, context, and content. In S. Kranger and S. Ohman, editors, *Philosophy and Grammar*, pages 79–100. Reidel, 1980.

David Lewis. Scorekeeping in a language game. In *Philosophical Papers, Vol.I*. Oxford University Press, 1983.

Stanley Peters. A truth-conditional formulation of Karttunen's account of presupposition. *Synthese*, 40:301–316, 1979.

Daniel Rothschild. Making dynamics semantics explanatory. manuscript, Columbia University, 2008a.

Daniel Rothschild. Presupposition projection and logical equivalence. *Philosophical Perspectives*, 22(1):473–497, 2008b.

Daniel Rothschild and Seth Yalcin. On the dynamics of conversation. *Noûs*, 2015.

Daniel Rothschild and Seth Yalcin. Three notions of dynamicness in language. *Linguistics and Philosophy*, 39(4):333–355, 2016.

Philippe Schlenker. Be articulate: A pragmatic theory of presupposition projection. *Theoretical Linguistics*, 34(3):157–212, 2008.

Philippe Schlenker. Local contexts. *Semantics and Pragmatics*, 2(3):1–78, 2009.

Philippe Schlenker. Donkey anaphora: the view from sign language (ASL and LSF). *Linguistics and Philosophy*, 34(4):341–395, 2011.

Scott Soames. Presuppositions. In D. Gabbay and F. Guenther, editors, *Handbook of Philosophical Logic*, volume IV, pages 553–616. Dordrecht, 1989.

Robert Stalnaker. Pragmatics. *Synthese*, 22:272–289, 1970.

Robert Stalnaker. On the representation of context. *Journal of Logic, Language and Information*, 7(1):3–19, 1998.

Matthew Stone. *Or* and anaphora. In Chris Barker and David Dowty, editors, *SALT 2*, page ß, 1992.

Rob van der Sandt. Presuppostion projection as anaphora resolution. *Journal of Semantics*, 9 (4):333–377, 1992.

Kai von Fintel. *Restrictions on Quantifier Domains*. PhD thesis, University of Massachusetts, Amherst, 1994.

Tests of scale structure theory in dimensional and multidimensional adjectives

Galit Weidman Sassoon
[1] Bar Ilan University, Ramat Gan, Israel.
galitadar@gmail.com

Abstract

This paper addresses the need to pay attention to the multiplicity of possible interpretations of adjectives when applying to them the standard tests of scale structure and standard (Kennedy & McNally 2005). In particular, the paper considers simple-dimensional, complex-dimensional, and multidimensional interpretations of multidimensional adjectives (Sassoon in progress).

1 Introduction

Standard theories of gradability associate gradable adjectives with a single scalar dimension per context, like height or health (Bierwisch, 1989; Kennedy, 2007; Kennedy & McNally, 2005; Rotstein & Winter, 2004). Often, the dimension is used to compose a relation between individuals and degrees – a denotation at type <d, <e,t>>. For example, *tall* denotes a relation between degrees d and entities x which are at least d tall. A null morpheme, *pos*, introduces a membership norm c called *standard* into the logical form and truth conditions of positive forms. For example, (1) is true iff Ann is at least as tall as the norm (von Stechow, 2007).

(1) Ann is pos_c tall.

Degrees can reflect a single measurement, like height (a basic dimension), or the output of a function over degrees in a set of measurements, like f_F with weights w_F $f(f_{F1}...f_{Fn}, w_{F1}...w_{Fn})$ (a complex dimension; Bylinina, 2013; Kennedy, 2013; McNally & Stojanovic, 2014; Umbach, 2016; Solt, 2018). For example, the optimism of an entity x can be modeled as the *weighted sum* $\mathbf{\sum w_F f_F(x)}$ of x's degrees in various measurements, and x's health can be modeled as a *weighted product* $\mathbf{\Pi w_F f_F(x)}$. Weighted products capture the intuition that, for example, any life-threatening disease reduces one's average health below any plausible standard, no matter how healthy one is otherwise. When degrees are multiplied, a low degree in a single dimension strongly reduces the overall product (for example, $1 \cdot ... \cdot 1 \cdot 0.5 = 0.5$) and thus reduces the classification probability. In contrast, when degrees are added, a few low degrees hardly affect the overall sum (for example, $1 + ... + 1 + 0.5$ is almost the maximal sum possible). This is useful to model cases in which the contributions of the different dimensions are independent, as is characteristic of traits like *optimistic* (Murphy 2002; Pothos & Wills 2011).

Indeed, recent work adopts the intuitive view that speakers weigh the dimensions of multidimensional adjectives by importance (Kennedy 2013; McNally & Stojanovic, 2014), and sum up the degree of the entities in those dimensions factored by their weights (Gärdenfors 2004; Bylinina, 2013; Solt, 2018). For example, Kennedy (2013) argues for uncertainty about how the dimensions involved in standard calculation are weighted in different situations. Kennedy thereby explains faultless disagreement in the presence of multiple dimension, namely, the fact that no side can be proven wrong when speakers disagree about whether the application of an adjective like *typical, beautiful* or *safe* to a given entity is truthful or not. Furthermore, as illustrated above, the cognitive literature often uses averaging functions to collapse the set of dimensional degrees and weights into a single degree. While more complex functions than averaging are also possible, this does not affect the arguments in this paper.

Thus, for example, on any of its **simple dimensional interpretations**, *healthy*, denotes a relation, R_F, between entities x and their levels d of health with respect to a contextually given dimension F (e.g., cholesterol, diabetics, flu or chickenpox; Bartsch 1984; Kennedy 2013). By contrast, on its **complex dimensional interpretation**, *healthy* denotes a relation, $R'_{healthy}$, between entities x and their 'averaged' health levels d (e.g., the weighted product, $\mathbf{\Pi w_F f_F(x)}$, of the degrees of entities x in the contextually relevant health indices F). With a maximum standard for *healthy* and a minimum standard for *sick*, entities are predicted to count as *pos healthy* iff they are maximally healthy in every respect, and as *pos sick* otherwise (Bylinina 2013).

However, besides adjectives whose scales are based on simple or complex dimensions of the sort illustrated above, Bartsch & Vennemann (1972) and Bartsch (1984) represented certain adjectives as multidimensional. Sassoon (in progress) and Sassoon & Fadlon (2017) argued that these adjectives have, in addition to any dimensional interpretation, also a multidimensional interpretation, where the scale is based on dimension counting.

For example, on its **multidimensional interpretation**, *healthy* denotes a relation, $R_{healthy}$, between entities x and the number n of dimensions with respect to which they are healthy. In this interpretation, the standard of *healthy* standard represents the minimal number of dimensions with respect to which entities have to be healthy in order to count as *pos healthy*. By virtue of a maximum standard, *pos healthy* conveys being healthy in every respect. *Sick* denotes a relation, R_{sick}, between entities x and the number n of dimensions with respect to which they are sick, and by virtue of a minimum standard, *pos sick* conveys being sick in at least one respect.

The multidimensional interpretation of an adjective differs from its dimensional and complex-dimensional interpretations in—often subtle but—important respects (see discussion in Sassoon in progress). The default multidimensional interpretation of a positive adjective like *safe*, for example, ('safe in every respect') asymmetrically entails any interpretation based on a specific dimension (for example, frequency of cases of robbery or rape). Evidence for its presence is the intuition that a neighborhood can be considered *not pos safe* even if the only piece of information available is that some big enough danger exists, namely there is a respect in which the neighborhood is not safe. No further information is needed about the nature of that danger (the respect being violated). This much information does not suffice to falsify an interpretation of *pos safe* based on a particular dimension like robbery.

By contrast, the multidimensional interpretation of *pos safe* (safe in every respect) is asymmetrically entailed by any **maximum-standard** complex (e.g., averaging-based) interpretation of *pos safe*. The two interpretations differ because a neighborhood which is not maximally safe in every respect (namely, not *pos safe* in the complex-dimension sense) can still be safe in every respect (namely, *pos safe* in the multidimensional sense). Evidence for the multidimensional interpretation comes from the intuition that a neighborhood can be considered *pos safe* even when some degree of danger of some sort exists, for example, robbery occurs but rarely enough that the neighborhood counts as safe in this respect. In general, a multidimensional interpretation is sensitive to the standards of the dimensions and does not necessitate maximal standards (relative adjectives may well constitute dimensions). By contrast, a complex dimensional interpretation is not sensitive to the standards of the dimensions, and, with a maximum standard reduces to a quantificational interpretation like "adjective in every respect", only assuming the dimensions have a maximum standard too.

In sum, intuitively, in some contexts, a neighborhood is considered (*pos*/*perfectly*) *safe* iff the neighborhood is safe in every respect, and *not* (*pos*/*perfectly*) *safe* otherwise. No account in terms of a unique dimensional interpretation (either simple or complex) captures both these truth condition and falsity condition simultaneously (Sassoon, in progress).

Section 2 briefly reviews some of the motivations for dimension-counting interpretations, and clarifies the distinctions between them and other interpretations of adjectives that involve counting scales (Sassoon, in progress). This multiplicity of interpretations gives rise to a need to make the standard tests of scale structure of adjectives more precise. In particular, section 3 reviews some of the scale structure tests. Some tests are based on judgments of inference patterns or contradictions between sentences with a given adjective. While judging whether an inference follows or a contradiction holds, it is important to control for the type of interpretation of each ambiguous or context-sensitive word in the premises and conclusions or in the potentially contradicting sentences. Section 3 suggests that to better understand the results of the tests of scale-structure theory in the presence of counting-based (quantificational) interpretations, the application of a supplementary test, based on exceptive phrases, is needed.

2 Motivations for the dimension-counting hypothesis

Exception phrases indicate universal generalizations as opposed to existence statements, as shown in the contrast in examples (3a,b) vs. (3c,d) (Hoeksema, 1995; Moltmann, 1995; von Fintel, 1994).

(1) a. Everyone arrived except for Mary.
b. No one arrived except for Mary.
c. #Someone arrived except for Mary.
d. #Not everyone arrived except for Mary.

Thus, the higher acceptability of exception phrases in examples like (4a) rather than (4c) seems to stem from a higher tendency to interpret positive forms of positive adjectives like *healthy*

as involving universal quantification over dimensions. In addition, the higher acceptability of exception phrases in examples like (4b) than those like (4d) seems to stem from a higher tendency to interpret positive forms of negative antonyms like *sick* as involving existential quantification over dimensions. Negation reverses the quantificational force, resulting in universal quantification in (4b) and existential quantification in (4d) (Hoeksema, 1995).

(2) a. Mary is **healthy** except for high cholesterol (ch)
b. Mary is **not sick** except for the flu
c.# Mary is **sick** except for normal cholesterol (ch)
d.# Mary is **not healthy** except for (normal) cholesterol (ch)

Judgment studies support the acceptability contrasts indicated in (2) (see review in Sassoon, in progress) and corpus studies reveal distributional patterns reflecting these judgments. Sassoon (2013) considered 1300 naturally occurring examples of the form 'Adj. except' with 8 antonym pairs in positive vs. negated contexts. Positive adjectives manifested mainly interpretations involving universal quantification over dimensions, while negative adjectives manifested mainly interpretations involving existential quantification.

Sassoon (in progress) argued that the basis for these trends is a scale based on dimension counting with a tendency toward a maximum standard in positive adjectives, as opposed to a minimum standard in negative ones. One motivation for this proposal was the observation that comparison constructions may have an interpretation in which dimension-cardinalities are directly compared. For instance, example (3), in addition to having an **access reading**, as in (3a), and a **quantificational reading**, as in (3b), can also have a **dimension-counting reading**, as in (3c).

(3) Ann is more successful than Bill.
a. Ann is more successful than Bill is in some **salient respect** (their math studies).
b. Ann is more successful than Bill is **in n-many (for example, most) respects**.
c. Ann is pos successful **in more respects** than Bill is.

In the dimension-counting proposal, *successful* denotes the dimension-counting relation that holds between degrees d and entities who are successful in at least d respects. Thus, comparisons like (3) can convey that there is a number d, such that Ann is successful in at least d respects, while Bill isn't, namely reading (3c). In addition, in this proposal, *(un)successful* denotes the dimension-counting relation that holds between degrees d and entities who are (un)successful in at least d respects. Thus, (4) is correctly predicted to have a dimension-counting interpretation, conveying that there are more dimensions in which Bill is successful than there are dimensions in which he is unsuccessful (Sassoon, in progress).

(4) Bill is more successful than unsuccessful.

In (5) (from 2009's academic section of the corpus of contemporary American English, Davis 2010), the number of dimensions in which reading and spelling are alike seem to compare to the number of dimensions in which they are different. The contextually supplied dimensions are language skills. Reading and spelling count as similar with respect to a given skill if both require it. Thus, reading and spelling count here as similar in the multidimensional sense of the dimension-counting proposal because (and to the extent that) they require the same skills.

(5) "Reading and spelling require the same language skills (Moats, 2005), have a strong correlation (Ehri, 2000), and support the development of each other (Snow, Griffin, &; Burns, 2005). **Reading and spelling are more similar than different**..."

Furthermore, intuitively, degree modified adjectives like, for example, *{perfectly, mostly, very, somewhat} happy,* may contribute information about the number of dimensions (e.g., all, most, many, some, respectively) whose norms their argument exceeds. Again, dimension-counting scales predict the availability of such readings. In sum, the dimension-counting account captures the wider set of interpretations of positive, comparative and degree-modified forms of multidimensional adjectives as compared with dimensional adjectives (Sassoon, in progress).

Moreover, while dimensional adjectives by definition do not have readings involving quantification over dimensions, they may have readings involving quantification over other types of objects. Such readings can also be diagnosed using exception phrases. For example, we can describe a crowded classroom using (6a), but we cannot describe a sick child using (6b).

(6) a. The classroom is full/empty except for one chair.
b. #The child is warm expect for one degree.

Arguably, these judgments stem from the tendency of *full* towards interpretations with a maximum standard as opposed *warm*, which does not have this tendency. In the context of the utterance in (6a), *full* is associated with a chair-counting scale ranging between 0 and the maximum number of chairs in the given classroom. Thus, *full* denotes the relation between degrees d and locations containing chairs x, which holds iff the number of occupied chairs in x is d. Since *full* tends toward a maximum standard, *pos full* truly apply to a location x iff every chair in x is occupied.

By contrast, *warm* in a context suitable for the utterance in (6b), has a conventional mid-scale standard of 36˚. Thus, *The child is pos warm* conveys that the child's temperature is warmer than 36˚. Since these truth conditions do not reduce to universal quantification over temperatures, an exceptive is not licensed. Again, we see a connection between exceptive licensing and maximal standards. Thus, exceptive licensing can form a test for maximal standards of cardinality scales, including dimension-counting or other object-counting scales.

In fact, Yoon (1994) argued for adjectival readings involving universal and existential quantification over individuals or subparts. Many of Yoon's examples are also multidimensional adjectives. For example, for a table or table part to be dirty it has to be dusty, stained, oily, crumby, or dirty in some other way, while for it to be clean it has to be clean in every way. Thus, positive forms like (7a,b) may involve quantification over parts, dimensions, or both.

(7) a. The tables are clean/not dirty (except for the one on the left, which is dusty).
b. The table is clean/not dirty (except for one part, which is slightly dusty).

The evaluation of whether a table/part is, e.g., *rather dirty*, *fairly clean,* or *cleaner* than another table/part depends on what precisely is being counted.

With this in hand, we move on to scale structure theory's tests.

3 Standard types predicted by the different hypotheses

Gradable predicates divide by whether their unique scale has a minimum, a maximum, both or neither, and whether their standard is identified with the scale maximum, minimum, or neither (Kennedy, 2007; Kennedy & McNally, 2005; Rotstein & Winter, 2004; Van Rooij, 2010; Syrett, 2007). In relative adjectives, like *interesting,* the standard is context relative, while in absolute adjectives the standard is a lexicalized scale-endpoint. For instance, one indication that *clean* has a maximum standard is the intuition that (8a), unlike (8b), is a contradiction, as the symbol $\#_C$ indicates. The source of the contradiction in (8a) is the inference in (9a). The consistency of (8b) correlates with the inference failure indicated in (9b).

(8) a. $\#_C$This table is clean, but that one is cleaner.
b. This table is beautiful, but that one is more beautiful.
(9) a. $\#_C$That table is cleaner than this one.
=> This table is not maximally clean. => This table is not clean.
b. That table is more interesting than this one.
=> This table is not (maximally) interesting. $\not\Rightarrow$ This table is not interesting.

An indication that *different* has a minimum-standard is the intuition that (10a) is a contradiction, unlike (10b). The source of the contradiction in (10a) is the inference in (11a). The consistency of (10b) correlates with the inference failure in (11b).

(10) a. $\#_C$This chair is not different from mine, but is more different than that one is.
b. This chair is not beautiful, but is more beautiful than that one is.
(11) a. $\#_C$This chair is more different from mine than that one is.
=> This chair is at least minimally different from mine.
=> This chair is different from mine.
b. This chair is more interesting than that one is.
=> This chair is at least minimally interesting. $\not\Rightarrow$ This chair is interesting.

These tests of standard type yield clear results in adjectives that are **not** multidimensional, like those used in much of the experimental work on scale structure theory (e.g., Syrett, Kennedy, & Lidz, 2009). This work addressed mostly basic scales of dimensional adjectives like

straight-bent, *full-empty* and *transparent-opaque*. To illustrate, examples (12a,b) are clearly inconsistent, indicating a maximum- and minimum-standard for *straight* and *bent*, respectively.

(12) a. $\#_C$ This rod is straight, but that one is more straight.
b. $\#_C$ This rod is not bent, but is more bent than this one.

However, when the interpretation of adjectives involves quantification over dimensions, the test results might be affected by the force of the quantifier over dimensions. In particular, the tests don't yield clear results when the default standards of the multidimensional interpretation, dimensional interpretation, and dimensions' interpretations are not identified with the same scale point. To illustrate, the tests demonstrate that the interpretations of *similar* or *familiar* are typically associated with **midpoint or minimum-standards**. The consistency of (13a,b) suggests that their standard is not identified with the scale-maximum. The fact that (13c,d) is judged as inconsistent, suggests that the standard is identified with the scale minimum, while them being judged as consistent may suggest a midpoint standard.

(13) a. This version is similar to the original draft, but that one is more similar.
b. This fruit looks familiar, but that one looks more familiar.
c. $?_C$ This version is not similar to the original, but it is more similar than that one.
d. $?_C$ This fruit does not look familiar, but it looks more familiar than that one.

However, such an application of the tests ignores the fact that *similar* and *familiar* have multidimensional interpretations. For example, the paper versions talked about in (13a) can be similar or not in font type, font size, line spacing, length, wording, topics, depth, or strength of argumentation. Thus, (13a) can relate to font size only, to a uniform complex dimension based on weighted summing over different respects, or to a scale based on counting different respects in which the two versions are similar. Similarly, a fruit can be familiar or not with respect to shape, size, color, taste, having seeds or not, having stripes or not, or serving certain purposes or not. Thus, again, *familiar* as applied to fruits has a variety of interpretations.

Moreover, we have shown that the exception-phrase tests of universality and existentiality over dimensions illustrated in (2) indicate whether a multidimensional interpretation tends to have a maximum or minimum standard. The corpus results reviewed in section 2 suggest that positive forms of positive adjectives often involve universal quantification over dimensions. In particular, the distribution of exception phrases suggests that *pos familiar* often conveys being familiar in all relevant respects. Since, according to the standard theory, the test in (13b) indicates that *familiar* **does not have a maximum standard**, the proposal that *familiar* is associated with a complex (e.g., averaging-based) dimension with a minimum or mid-point standard cannot explain *familiar*'s tendency toward universal quantification over dimensions. Following this proposal, fruits are predicted to count as familiar iff they are familiar to at least some non-maximal degree in a single (basic or complex) dimensional scale. Thus, it does not follow that familiar fruits are familiar in every respect.

However, this tendency can be captured assuming a multidimensional (dimension-counting) interpretation for this adjective that does involve **a maximum standard**. Given the empirical

generalization that positive adjectives tend to be universal over their dimensions (see section 2), being positive, the default interpretation of *pos familiar* is familiar in every respect. Given this universal interpretation, exceptives are correctly predicted to be licensed.

In fact, when the adjectives in Sassoon's (2013) corpus study were divided into adjectives with default maximum and minimum standards by the standard tests of scale structure theory, the two adjective-sets did not differ significantly in their universality vs. existentiality index (exceptive frequencies). The reason was precisely that some maximum standard adjectives were negative and thus more existential than universal over their dimensions (like *unfamiliar*), while some midpoint or minimum standard adjectives were positive and thus more universal than existential over their dimensions (like *familiar*).

Hence, the proposal that multidimensional adjectives have an interpretation based on dimension-counting, together with the generalization that multidimensional interpretations tend toward a maximum standard in positive adjectives and a minimum standard in negative adjectives, is needed to capture the corpus data and judgments in (13). Based on the proposal and generalization, the multidimensional truth conditions of positive forms of, for example, *familiar*, require that for **all** familiarity respects F, the individual talked about would be at least **somewhat** familiar with respect to F. The consistency of (13b) follows, because other individuals may be even more familiar than the fruit talked about. They may, for instance, be **very** familiar in the given respects.

As for (13d), it is only predicted to be inconsistent assuming the standard hypothesis that *familiar* is associated with a complex (e.g., averaging-based) dimension, by which *not pos familiar* means not being even a bit familiar (having a zero degree of familiarity). By contrast, assuming a multidimensional interpretation along the dimension-counting proposal, *not pos familiar* is consistent with being a bit familiar in some respects or others, and therefore more familiar than individuals who are in no way even a bit familiar. Such predictions are consistent with the shaky status of speakers' judgments about (13c,d).

Moreover, in the corpus investigated by Sassoon (2013), almost all the adjectives that were usually universal over their dimensions, like *typical*, admitted some existential uses and vice versa, regardless of the type of standard typically associated with them in scale-structure theory. In those cases, a complex dimensional interpretation with the standard identified by the scale structure theory tests cannot provide a sufficient account for the quantification over dimensions, and a multidimensional interpretation based on dimension-counting is needed to do the job. Dimension-counting scales seem to usually be closed on both sides ranging between 0 and the cardinality of the entire dimension set. Thus, these standard shifts are possible.

Moreover, often speaker judgments are not conclusive when applying scale structure theory's tests even with adjectives like *safe, clean* or *healthy,* which are usually maximum-standard. This happens because they may admit mid-point relative interpretations as well. Thus, speakers diverge on whether the examples in (14) are inconsistent as the literature assumes (Kennedy, 2007; Kennedy & McNally, 2005; Rotstein & Winter, 2004), or not.

(14) ?$_{\mathrm{C}}$ This is {clean, safe, healthy}, but that is (even) {cleaner, safer, healthier}.

Context relativity may pervade the interpretation of absolute multidimensional adjectives either through a shift to non-absolute standards (e.g., *safe* may convey being safe in most respects), or through a context-relative dimension (e.g., *safe with respect to robbery*). Then there are no entailment relations between the multidimensional and complex-dimension interpretations (for context effects in absolute adjectives see Bierwisch 1989; McNally 2011).

To conclude, this paper proposes an account for cases in which speakers' judgments on the tests of scale structure theory are less conclusive than expected assuming simple or complex dimensional interpretations of adjectives. The tests provide important information about the types of standards of adjectives, but they must be used with caution in order not to confound the types of standard of the different possible interpretations of each adjective, including, in particular, interpretations involving quantification over dimensions. When these exist, the test results might be affected both by the standards of the dimensions and by the force of quantification over dimensions (the standard of the multidimensional interpretation), as indicated by the exceptive phrase tests. These tests are needed to supplement judgments.

Sassoon (in progress) shows that this conclusion applies to additional tests of scale structure theory. For example, in this theory, absolute modifiers like *perfectly* select adjectives whose scale has a maximum (Rotstein & Winter 2004; Kennedy, 2007; Kennedy & McNally, 2005; Syrett, 2007). By contrast, absolute modifiers like *somewhat* or *slightly* select adjectives whose scale includes a minimum. However, Sassoon (in progress) illustrates that degree modified multidimensional adjectives have dimension-counting interpretations (see section 2). Moreover, in Sassoon's (2013) study, the universality score of each adjective (as given by the frequencies of exceptive-modification of its occurrences) strongly correlated with the frequency of its modification by *perfectly* ($r = 0.7$). Universality did not correspond with having a maximum standard by the tests of scale structure theory. Overall, this supports an account of *perfectly* as selecting multidimensional adjectives that are universal over their dimensions, even when their simple- or complex-dimensional interpretations have relative or minimum standards according to the tests of scale structure theory (e.g., *beautiful* or *similar*). Again the exceptive tests are needed to supplement scale structure theory's tests.

References

Bartsch, Renate & Theo Vennemann. 1972. The grammar of relative adjectives and comparison. *Linguistische Berichte 20*: 19-32.

Bartsch, Renate. 1984. The structure of word meanings: Polysemy, Metaphor, Metonymy. Fred Landman & Frank Veltman (eds.), *Varieties of Formal Semantics*: 25–54. Foris Publications. Dordrecht.

Bierwisch, Manfred. (1989). The semantics of gradation. In Manfred Bierwisch & Ewald Lang (eds.), *Dimensional adjectives: Grammatical structure and conceptual interpretation*, 71–261. Springer-Verlag.

Bylinina, Lisa. 2013. *The grammar of standards.* Doctoral dissertation, Utrecht University.

Gardenfors, Peter. (2004). *Conceptual Spaces—The Geometry of Thought*. MIT Press. Cambridge, Massachusetts.

Hoeksema, Jack. (1995). The semantics of exception phrases. In J. van Eijck & J. van der Does (eds.) *Quantifiers, Logic, and Language*, 145-177. CSLI. Stanford.

Kennedy, Chris. (2007). Vagueness and grammar: The semantics of relative and absolute gradable adjectives. *Linguistics and philosophy, 30*(1), 1-45.

Kennedy, Chris. (2013). Two sources of subjectivity: Qualitative assessment and dimensional uncertainty. *Inquiry, 56*(2-3), 258-277.

Kennedy, Chris., & Louise McNally. 2005. Scale structure, degree modification, and the semantics of gradable predicates. *Language*, 345-381.

Lasersohn, Peter. (1999). Pragmatic halos. *Language*, 522-551.

McNally, Louise. (2011). *The relative role of property type and scale structure in explaining the behavior of gradable adjectives.* In Rick Nouwen, Robert van Rooij, Uli Sauerland & Hans-Christian Schmitz (eds.) *Vagueness in Communication 6517*: 151-168. Springer, Berlin.

McNally, Louise & Isidora Stojanovic. (2015). Aesthetic adjectives. In James Young. (ed.) Th*e semantics of aesthetic judgment*, Oxford: Oxford University Press.

Moltmann, Frederique. (1995). Exception sentences and polyadic quantification. *Linguistics and philosophy, 18*(3), 223-280.

Murphy, Gregory. (2002). *The Big Book of Concepts*. The MIT Press. Cambridge, MA.

Pothos, Emmanuel M. and Wills, Andy J. (2011). *Formal Approaches in Categorization*. Cambridge University Press, Cambridge, UK.

Rotstein, Carmen & Yoad Winter. (2004). Total adjectives vs. partial adjectives: Scale structure and higher-order modifiers. *Natural Language Semantics, 12*(3), 259-288.

Sassoon, Galit W. (In progress). Multidimensionality in the grammar of gradability. In review process. Bar Ilan University.

Sassoon, Galit W. (2013). A typology of multidimensional adjectives. *Journal of Semantics* 30: 335-380.

Sassoon, Galit W. & Julie Fadlon. (2017). The role of dimensions in classification under predicates predicts their status in degree constructions. *Glossa: a journal of general linguistics*, 2(1), 42. DOI: http://doi.org/10.5334/gjgl.155

Solt, Stephanie. (2018). Multidimensionality, subjectivity and scales: experimental evidence. In Elena Castroviejo, Louise McNally & Galit Weidman Sassoon (Eds.) *The Semantics of Gradability, Vagueness, and Scale Structure -Experimental Perspectives.* Language, Cognition, and Mind, Springer: Switzerland.

Syrett, Kristen, Chris Kennedy & Jef Lidz. (2009). Meaning and context in children's understanding of gradable adjectives. *Journal of semantics*, 27: 1-35.

Syrett, Kristen. (2007). *Learning about the structure of scales: Adverbial modification and the acquisition of the semantics of gradable adjectives*. Doctoral dissertation. Northwestern University, Department of Linguistics.

Umbach, Carla. (2016). Evaluative propositions and subjective judgments. In Meier, Cecil & Jannet van Wijnbergen-Huitink. (Eds.) S*ubjective meaning: Alternatives to Relativism*. Walter de Gruyter GmbH and Co KG.

van Rooij, Robert. (2010). Measurement and interadjective comparisons. *Journal of semantics* 28(3): 335-358.

von Fintel, Kai. 1994. *Restrictions on quantifier domains.* Doctoral dissertation, University of Massachusetts, Amherst.

von Stechow, Arnim. (2007). The temporal adjectices früh (er)/spät (er) and the semantics of the positve. In Anastasia Giannakidou and Monika Rathert. (Eds.) *Quantification, definiteness, and nominalization*, 214-233. Oxford: Oxford University Press.

Yoon, Youngeun. (1996). Total and partial predicates and the weak and strong interpretations. *Natural Language Semantics, 4*(3), 217-236.

Disjunctive Antecedents for Causal Models

Mario Günther

LMU Munich
mario.guenther@campus.lmu.de

Abstract

Sartorio [4] argues convincingly that disjunctive causes exist. To treat disjunctive causes within Halpern and Pearl [2]'s framework of causal models, we extend their causal model semantics by disjunctive antecedents and propose a refinement of their definition of actual causation.

1 Introduction

Halpern and Pearl [2] define actual causation based on a causal model semantics of conditionals. The semantics is restricted to antecedents that do not contain disjunctions. "We might consider generalizing further to allow disjunctive causes", so Halpern and Pearl [2, p. 853], but they discard the idea, because there be "no truly disjunctive causes once all the relevant facts are known".

In contrast, Sartorio [4] argues for the existence of disjunctive causes by putting forward a switching scenario, in which all the relevant facts are known. Sartorio's Switch provides motivation to extend Halpern and Pearl [2]'s causal model semantics and definition of actual causation to be applicable to causes that have a particular disjunctive form. Accordingly, we lift the restriction of causal models to non-disjunctive antecedents such that we can express arbitrary Boolean combinations in a conditional's antecedent.

In Section 2, we translate Sartorio's Switch in a causal model. En passant we introduce Halpern and Pearl [2]'s causal model semantics and definition of actual causation. In Section 3, we extend Halpern and Pearl [2]'s causal model semantics by antecedents having a disjunctive form. This allows us to refine Halpern and Pearl's definition of actual causation such that it captures disjunctive causes of the type found in Sartorio's Switch.

2 Sartorio's Switch and Causal Models

Sartorio [4] argues for the existence of disjunctive causes. She invokes roughly the following scenario to back up her claim.

Example 1. **Sartorio's Switch** (Sartorio [4, p. 523–528])
Suppose a train is running on a track onto which a person is tied. Although there is a switch determining on which of two tracks the train continues, the tracks reconverge before the place, where the person is captivated. Now, Sartorio adds details to this typical switching scenario. A person, called Flipper, flips the switch such that the train continues on the left track. Moreover, there is construction work carried out on the right track. Another person, called Reconnecter, reconnects the right track before the train would have arrived in case Flipper hadn't flipped the switch. The train travels on the left track and kills the trapped person.

Sartorio proposes that the disjunction 'Flipper flips the switch and/or Reconnecter reconnects' is the actual cause of the person's death, while both individually 'Flipper flips the switch' and 'Reconnecter reconnects' are not actual causes of the person's death.

In her judgment, she complies with Lewis [3]'s simple counterfactual analysis of actual causation. 'Flipper flips the switch' (and 'Reconnecter reconnects') is not an actual cause of the person's death. For, if it were not the case that 'Flipper flips the switch' (or 'Reconnecter reconnects' respectively), the person would die nevertheless. Additionally, the conjunction 'Flipper flips the switch and Reconnecter reconnects' is no actual cause of the person's death. For, if it were not the case, the person might die nevertheless, viz. in case one of Flipper and Reconnecter does what they do. However, the disjunction 'Flipper flips the switch or Reconnecter reconnects' is an actual cause of the person's death. For, if it were not the case, the person would not die. Sartorio [4, p. 530] confirms that "the death happened because *at least one of them* did what they did."

Sartorio [4] makes the intuition strong that Flipper's redirection is not a cause given that there was an alternative route, even if that route is never actualized. She thinks that "the mere fact that there was an alternative route is sufficient to rob the event of the redirection of its causal powers." (p. 532) Accordingly, Flipper's redirection to the left track renders Reconnecters reconnection of the right track causally inefficacious, and, conversely, the reconnection renders the redirection causally inefficacious. The core of her reasoning goes as follows: "If either event had happened without the other, then that event would have been causally efficacious [...]. But, when both events happen, they deprive each other of causal efficacy." (p. 531) However, so argues Sartorio, the outcome must still depend on the existence of some viable causally efficacious path. Hence, the disjunctive fact that at least one path was causally efficacious is the cause of the outcome.

We translate now Sartorio's Switch in a causal model and check which formulas qualify as actual causes according to Halpern and Pearl [2]'s definition of actual causation.

2.1 Halpern and Pearl's Causal Model Semantics

Halpern and Pearl [2, pp. 851-852]'s causal model semantics of conditionals is defined with respect to a causal model over a signature.

Definition 1. Signature
A signature $\mathcal{S}$ is a triple $\mathcal{S} = \langle \mathcal{U}, \mathcal{V}, \mathcal{R} \rangle$, where $\mathcal{U}$ is a finite set of exogenous variables, $\mathcal{V}$ is a finite set of endogenous variables, and $\mathcal{R}$ maps any variable $Y \in \mathcal{U} \cup \mathcal{V}$ on a non-empty (but finite) set $\mathcal{R}(Y)$ of possible values for Y.

Definition 2. Causal Model
A causal model over signature $\mathcal{S}$ is a tuple $M = \langle \mathcal{S}, \mathcal{F} \rangle$, where $\mathcal{F}$ maps each endogenous variable $X \in \mathcal{V}$ on a function $F_X : (\times_{U \in \mathcal{U}} \mathcal{R}(U)) \times (\times_{Y \in \mathcal{V} \setminus \{X\}} \mathcal{R}(Y)) \mapsto \mathcal{R}(X)$.

The mapping $\mathcal{F}$ defines a set of (modifiable) structural equations modeling the causal influence of exogenous and endogenous variables on other endogenous variables. The function F_X determines the value of $X \in \mathcal{V}$ given the values of all the other variables in $\mathcal{U} \cup \mathcal{V}$. Note that $\mathcal{F}$ defines no structural equation for any exogenous variable $U \in \mathcal{U}$.

Intuitively, a simple conditional $[Y = y]X = x$ is true in a causal model M given context $\vec{u} = u_1, ..., u_n$, if the intervention setting $Y = y$ results in the solution $X = x$ for the structural equations.[1] Such an intervention induces a submodel $M_{Y=y}$ of M.

[1] The solution is unique, because we consider only recursive causal models. We write $\vec{X}$ for a (finite) vector of variables $X_1, .., X_n$, and $\vec{x}$ for a (finite) vector of values $x_1, .., x_n$ of the variables. Hence, we abbreviate $X_1 = x_1, .., X_n = x_n$ by $\vec{X} = \vec{x}$. For simplicity, we do not properly distinguish between the vector and its set $\{\vec{X} = \vec{x}\}$.

Definition 3. Submodel
Let $M = \langle \mathcal{S}, \mathcal{F} \rangle$ be a causal model, $\vec{X}$ a (possibly empty) vector of variables in $\mathcal{V}$ and $\vec{x}, \vec{u}$ vectors of values for the variables in $\vec{X}, \vec{U}$. We call the causal model $M_{\vec{X}=\vec{x}} = \langle \mathcal{S}_{\vec{X}}, \mathcal{F}^{\vec{X}=\vec{x}} \rangle$ over signature $\mathcal{S}_{\vec{X}} = \langle \mathcal{U}, \mathcal{V} \setminus \vec{X}, \mathcal{R}\big|_{\mathcal{V}\setminus\vec{X}} \rangle$ a submodel of M. $\mathcal{F}^{\vec{X}=\vec{x}}$ maps each variable in $\mathcal{V} \setminus \vec{X}$ on a function $F_Y^{\vec{X}=\vec{x}}$ that corresponds to F_Y for the variables in $\mathcal{V} \setminus \vec{X}$ and sets the variables in $\vec{X}$ to $\vec{x}$.

We can describe the structure of Sartorio's Switch using a causal model including four binary variables:

- an exogenous variable T, where $T = 1$ if the train arrives and $T = 0$ otherwise;
- an endogenous variable F, where $F = 1$ if Flipper flips the switch and $F = 0$ otherwise;
- an endogenous variable R, where $R = 1$ if Reconnecter reconnects and $R = 0$ otherwise;
- an endogenous variable D, where $D = 1$ if the person dies and $D = 0$ otherwise.

Leaving the functions F_F, F_R and F_D implicit, the set of structural equations is given by:

- $F = T$
- $R = T$
- $D = max(F, R)$

In words, Flipper flips the switch ($F = 1$), if the train arrives ($T = 1$). Reconnecter reconnects ($R = 1$), if the train arrives. The person dies ($D = 1$), if at least one of $F = 1$ and $R = 1$ is the case. These recursive dependencies of the structural equations are depicted in Figure 1.

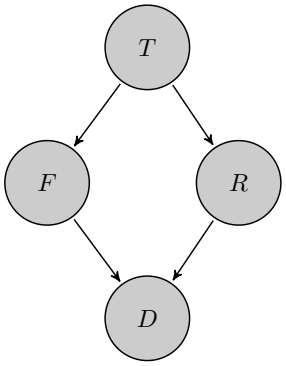


Figure 1: The causal network for Sartorio's Switch. The arrows represent the dependences of the structural equations.

To illustrate the causal model semantics, let us check whether or not the conditional $[F = 1]D = 1$ is true in the causal model M of Sartorio's Switch (given the context $t = 1$). Intuitively, the intervention that sets $F = 1$ induces a submodel $M_{F=1}$ of M. If the solution to the structural

equations of $M_{F=1}$ satisfies $D = 1$, then $[F = 1]D = 1$ is true in the causal model M under context $T = t$. In this case, we write $\langle M, t\rangle \models [F = 1]D = 1$.

In the scenario of Sartorio's Switch, $\langle M, t\rangle \models [F = 1]D = 1$ iff $\langle M_{F=1}, t\rangle \models D = 1$. The structural equations for the submodel $M_{F=1}$ are:

- $F = 1$
- $R = T$
- $D = max(F, R)$

We see that the solution to the structural equations of $M_{F=1}$ satisfies $D = 1$, and thus M satisfies the conditional $[F = 1]D = 1$ (given t). Notice the difference between the structural equation $F = T$ and $F = 1$: the former depends on T, whereas the latter does not. After the intervention that sets $F = 1$, the variable F is treated similar to an exogenous variable, i. e. it is assigned a value by its structural equation that does not depend on other (exogenous and/or endogenous parent) variables.[2] The structural equations for the variables in $\mathcal{V} \setminus \{F\}$ remain unchanged.

2.2 Halpern and Pearl's Definition of Actual Causation

The basic idea behind Halpern and Pearl [2]'s definition is to extend Lewis's notion of causal dependence to a notion of contingent dependence. Lewis [3, p. 563] defines causal dependence between two occurring events C and E in terms of counterfactual dependence. E causally depends on C iff (i) C and E occurred, and (ii) the simple counterfactual criterion is satisfied: if C had not happened, E would not have happened. Furthermore, he identifies actual causation with the transitive closure of causal dependence. Hence, C is an actual cause of E iff there is a chain of causal dependencies from C to E. Halpern and Pearl extend this definition by (possibly non-actual) contingencies: C is an actual cause of E iff E causally depends on C *under certain contingencies*. Roughly, contingent dependence makes it possible that even if E does not counterfactually depend on C in the actual situation, E counterfactually depends on C under certain contingencies.[3]

Based on their causal model semantics for conditionals, Halpern and Pearl [2, p. 853] propose the following definition of actual causation.

Definition 4. Actual Causation
$\vec{X} = \vec{x}$ is an actual cause of ϕ in $\langle M, \vec{u}\rangle$ iff the following three conditions hold:

AC1. $\langle M, \vec{u}\rangle \models (\vec{X} = \vec{x}) \wedge \phi$.

AC2. There exists a partition $\langle \vec{Z}, \vec{W}\rangle$ of $\mathcal{V}$ with $\vec{X} \subseteq \vec{Z}$ and some setting $\langle \vec{x'}, \vec{w'}\rangle$ of the variables in $\langle \vec{X}, \vec{W}\rangle$ such that if $\langle M, \vec{u}\rangle \models Z = z^*$ for all $Z \in \vec{Z}$, then both of the following conditions hold:

(a) $\langle M, \vec{u}\rangle \models [\vec{X} = \vec{x'}, \vec{W} = \vec{w'}]\neg\phi$.

(b) $\langle M, \vec{u}\rangle \models [\vec{X} = \vec{x}, \vec{W'} = \vec{w'}, \vec{Z'} = \vec{z^*}]\phi$ for all subsets $\vec{W'}$ of $\vec{W}$ and all subsets $\vec{Z'}$ of $\vec{Z}$.

[2] Intuitively, we may think of a value assignment $X = x$ in model M by an intervention as overruling the structural equation in M.

[3] Note that Halpern and Pearl do not take the transitive clossure for their definition of actual causation. In contrast to Lewis's dictum, they think [2, p. 844] that causation is not always transitive.

AC3. $\vec{X}$ is minimal; no subset of $\vec{X}$ satisfies conditions AC1 and AC2.

AC1 requires both that the actual cause $\vec{X} = \vec{X}$ and its effect ϕ are true in the actual (contextualized) model. AC3 ensures that only the conjuncts of $\vec{X} = \vec{x}$ "essential" for changing ϕ in AC2(a) are part of a cause: "inessential elements are pruned." (Halpern and Pearl [2, p. 853]) As proven by Eiter and Lukasiewicz [1], AC3 implies that an actual cause is always a single conjunct of the form $X = x$, if the set of endogenous variables is finite.

To understand AC2, it is helpful to think of $\vec{X} = \vec{x}$ as the minimal set of conjuncts that qualifies as a cause of the effect ϕ, and to think of $\vec{Z} = \vec{z}$ as the active causal path(s) from $\vec{X}$ to ϕ.

AC2(a) is reminiscent of Lewis [3]'s simple counterfactual criterion: ϕ would be false, if it were not for $\vec{X} = \vec{x}$. The condition says that there is a setting $\vec{X} = \vec{x}$ changing ϕ to $\neg\phi$, if the variables not on the active causal path(s) take on certain values, i. e. $\vec{W} = \vec{w'}$. The difference to the counterfactual criterion is that ϕ's dependence on $\vec{X} = \vec{x}$ may be tested under certain contingencies $\vec{W} = \vec{w'}$, which are non-actual for $\vec{w'} \neq \vec{w}$. Note that those contingent tests allow to identify more causal relationships than the simple counterfactual criterion.

AC2(b) restricts the contingencies allowed to be considered. The idea is that any considered contingency does not affect the active causal path(s) with respect to $\vec{X} = \vec{x}$ and ϕ. In other words, AC2(b) guarantees that $\vec{X}$ alone is sufficient to change ϕ to $\neg\phi$. The setting of a contingency $\vec{W} = \vec{w'}$ only eliminates spurious side effects that may hide $\vec{X}'s$ effect. The idea behind AC2(b) is implemented as follows: (i) setting a contingency $\vec{W} = \vec{w'}$ leaves the causal path(s) unaffected by the condition that changing the values of any subset $\vec{W'}$ of $\vec{W}$ from the actual values $\vec{w}$ to the contingent values $\vec{w'}$ has no effect on ϕ's value. (ii) At the same time, changing the values of $\vec{W'}$ may alter the values of the variables in $\vec{Z}$, but this alteration has no effect on ϕ's value.

We apply now Halpern and Pearl [2]'s definition of actual causation to the causal model of Sartorio's Switch. The result is that each of $F = 1$ and $R = 1$ is an actual cause of $D = 1$. However, the conjunction $F = 1 \wedge R = 1$ and the disjunction $F = 1 \vee R = 1$ do not qualify as actual causes of $D = 1$.

We show that $F = 1$ is an actual cause of $D = 1$. (The argument for $R = 1$ is structurally the same as the causal model of Sartorio's Switch is symmetric with respect to F and R.) Let $\vec{Z} = \{F, D\}$, and so $\vec{W} = \{R\}$. The contingency $R = 0$ satisfies the two conditions of AC2: AC2(a) is satisfied, as setting $F = 0$ results in $D = 0$; AC2(b) is satisfied, as setting F back to 1 results in $D = 1$. The counterfactual contingency $R = 0$ is required to reveal the hidden dependence of D on F, or so argue Halpern and Pearl.

We show that $F = 1 \wedge R = 1$ is not an actual cause of $D = 1$ due to the minimality condition AC3. Let $\vec{Z} = \{F, R, D\}$, and so $\vec{W} = \emptyset$. AC2(a) is satisfied, as setting $F = 0 \wedge R = 0$ results in $D = 0$. AC2(b) is satisfied trivially. However, two subsets of $\vec{X} = \{F, R\}$ satisfy the two conditions of AC2 as well, viz. $\vec{X'} = \{F\}$ and $\vec{X} = \{R\}$. Therefore, $\vec{X} = \{F, R\}$ is not minimal and according to AC3 the conjunction $F = 1 \wedge R = 1$ is thus no actual cause of $D = 1$. Minimality is meant to strip "overspecific details from the cause." (Halpern and Pearl [2, p. 857])

The disjunction $F = 1 \vee R = 1$ does not qualify as actual cause of $D = 1$, simply because Halpern and Pearl [2]'s definition of actual causation does not admit causes in form of proper disjunctions, i. e. disjunctions having more than one disjunct. They do not "have a strong intuition as to the best way to deal with disjunction in the context of causality and believe that disallowing it is reasonably consistent with intuitions." (p. 858)

Sartorio [4, p. 530] observes that "there is no general motivation for believing that, when (if)

a disjunctive fact is a cause, at least one of its disjuncts must also be a cause." This observation stands in sharp contrast to Halpern and Pearl [2]'s definition of actual causation, according to which both disjuncts individually qualify as actual causes. In the next section, we first define disjunctive antecedents for Halpern and Pearl's causal model semantics; subsequently, we extend their definition of actual causation to cover disjunctive causes as found in Sartorio's Switch.

3 An Extension of Causal Model Semantics by Disjunctive Antecedents

Recall Sartorio's Switch of Section 2. Sartorio argues that the person tied to the tracks dies because at least one of Flipper and Reconnecter does what they do. Therefore, the disjunctive fact that at least one track or path was causally efficacious is the cause of the outcome. Moreover, if only one of Flipper's and Reconnecter's events would occur, their *disjunction* would be causally inefficacious, but the single occuring event would be causally efficacious. We identify here two necessary conditions under which there are disjunctive causes: (i) there are more than one potentially efficacious and actually occuring events on different paths ("two tracks"), and (ii) there is an event that switches the paths without being, intuitively, a cause of the outcome ("flipping the switch").

Let us consider Sartorio's Switch using the variables of our causal model. In her switching scenario, Sartorio maintains that $F = 1 \vee R = 1$ is an actual cause of $D = 1$. The disjunction means that $D = 1$ because at least one of $F = 1$ and $R = 1$. On closer inspection, using our identified necessary conditions for disjunctive causes, Sartorio's disjunction means: the actual case $F = 1$ and $R = 1$ results in $D = 1$ *and* the counterfactual case $F = 1$ and $R = 0$ results in $D = 1$ *and* the counterfactual case $F = 0$ and $R = 1$ results in $D = 1$. There are two reasons: (a) if $F = 1$ alone were not sufficient to result in $D = 1$, the disjunction $F = 1 \vee R = 1$ would not be the actual cause. (Mutatis mutandis for $R = 1$.) (b) Both of $F = 1$ and $R = 1$ need actually to be the case. In such a case, if one *or* the other is sufficient for the effect and both occur, then Sartorio judges the disjunction of both to be the cause. In this sense, Sartorio understands the disjunction $F = 1 \vee R = 1$ as a summary of two actually occuring events $F = 1$ and $R = 1$, whose actual co-occurrence robs them of their individual causal efficacy, and which would, individually, be actual causes.

Halpern and Pearl [2]'s causal model semantics does not allow to evaluate the conditional $[F = 1 \vee R = 1]D = 1$. The reason is that they do not allow for disjunctions in the antecedent, and so the submodel $M_{F=1 \vee R=1}$ is undefined. Moreover, the structural equation for D of Sartorio's Switch does apply to values of F and R, but it does not apply to a disjunction such as $F = 1 \vee R = 1$. Hence, the value for D is not determined by the disjunction. Next, we propose a conservative extension of Halpern and Pearl's causal model semantics that allows us to evaluate antecedents that are disjunctive in Sartorio's sense.

3.1 Evaluating Disjunctive Antecedents

As we have just observed, Sartorio's disjunctive causes of the form $A = a \vee B = b$ require that $A = a \wedge B = b$ actually obtain, and if one of $A = a$ or $B = b$ would obtain but not the other, the effect would still follow. We implement now this logic governing Sartorio's disjunctive causes by extending Halpern and Pearl [2]'s framework of causal models.

The idea behind evaluating a conditional with disjunctive antecedent is to check whether the consequent is true in *each* disjunctive situation of the antecedent. We say that a Sartorio

disjunction $A = a \vee B = b$ is satisfied if three possible situations are satisfied: (i) $A = a \wedge B = b$, (ii) $A = a \wedge B = \neg b$, and (iii) $A = \neg a \wedge B = b$. We refer to (i)-(iii) as the disjunctive situations or possibilities of the formula $A = a \wedge B = b$. Intuitively, each disjunctive situation corresponds to one intervention that sets the values for a non-disjunctive formula. The result is one submodel per disjunctive situation. The antecedent $[A = a \vee B = b]$, for example, does not correspond to a unique intervention, but rather to three interventions. Each of the interventions results in exactly one submodel. The intervention (i), for instance, results in the submodel $M_{A=a,B=b}$, in which A and B take the same values than in the actual contextualized model $\langle M, \vec{u}\rangle$ given $A = a \vee B = b$ is an actual disjunctive cause in Sartorio's sense.

To evaluate a conditional with disjunctive antecedent does – according to the outlined idea – not require to modify Halpern and Pearl [2, pp. 849–852]'s notion of a submodel. Rather, the evaluation requires to look at (possibly) more than one submodel, namely at exactly one submodel for each disjunctive situation. In general, we write ϕ_i, where $(1 \leq i \leq n)$, for the formula that expresses the i-th disjunctive situation of the formula ϕ (that contains only finitely many primitive events).[4]

For clarity, we define an extended causal language.

Definition 5. Extended Causal Language $\mathcal{L}$
The extended causal language $\mathcal{L}$ conains

- the two propositional constants $\top$ and $\bot$,
- a finite number of random variables $\vec{X} = X_1, ..., X_n$ associated with finite ranges $\mathcal{R}(X_1), ..., \mathcal{R}(X_n)$,
- the Boolean connectives $\wedge, \vee, \neg$ and the operator $[]$, and
- left and right parentheses.

A formula ϕ of $\mathcal{L}$ is well-formed iff ϕ has the form

- $X = x$ for $x \in \mathcal{R}(X)$ (primitive event);
- if $\phi, \psi \in \mathcal{L}$, then $\neg\phi, \phi \wedge \psi, \phi \vee \psi \in \mathcal{L}$ (Boolean combinations of primitive events);
- if $[]$ does not occur in $\phi, \psi \in \mathcal{L}$, then $[\phi]\psi \in \mathcal{L}$ (causal conditionals).

For the extended causal language, we define a valuation function. Recall that $\langle M, \vec{u}\rangle \models X = x$ is shorthand for $X = x$ is the solution to all of the structural equations in the recursive model M given context $\vec{u}$.

Definition 6. Valuation Function
A valuation function $v_{\langle M,\vec{u}\rangle}$ (abbreviated as v) is associated with any arbitrary model M and any arbitrary vector $\vec{u}$. $v_{\langle M,\vec{u}\rangle} : \mathcal{L} \mapsto \{1, 0\}$ assigns either 1 or 0 to all formulas of the extended causal language $\mathcal{L}$:

$$\text{(a)}\ \ v(X = x) = \begin{cases} 1, \text{ if } \langle M, \vec{u}\rangle \models X = x \\ 0, \text{ otherwise} \end{cases}$$

[4] Note that the number i of ϕ_i depends on the number d of disjunctions occurring in ϕ. In general, $i \leq 2^{d+1}-1$, i. e. there are at most $2^{d+1} - 1$ disjunctive situations of ϕ. When the disjuncts are mutually exclusive, there are less disjunctive situations, because some are impossible. Take for example $F = 1 \vee F = 0$ for the binary variable F. Here, there are only two disjunctive situations, because $F = 1 \wedge F = 0$ is impossible. For, if $v(F = 1) = 1$ then $v(F = 0) = 0$ and if $v(F = 0) = 1$ then $v(F = 1) = 0$.

(b) $v(\neg\phi) = 1$ iff $v(\phi) = 0$

(c) $v(\phi \wedge \psi) = 1$ iff $v(\phi) = 1$ and $v(\psi) = 1$

(d) $v(\phi \vee \psi) = 1$ iff $v(\phi) = 1$ or $v(\psi) = 1$

(e) $v([\phi]\psi) = \begin{cases} 1, \text{ if } v(\psi) = 1 \text{ in each } \langle M_{\phi_i}, \vec{u}\rangle \\ 0, \text{ otherwise} \end{cases}$

, where M_{ϕ_i} is a submodel of M such that $\langle M, \vec{u}\rangle \models [\phi_i]\psi$, and ϕ_i is a non-disjunctive formula expressing one disjunctive possibility of ϕ.

Clause (c) of the valuation function entails that $X_1 = x_1, .., X_n = x_n$ is the setting of the variables in the contextualized model $\langle M, \vec{u}\rangle$ iff $X_1 = x_1 \wedge ... \wedge X_n = x_n$ is true in $\langle M, \vec{u}\rangle$. Hence, a vector of primitive events $\vec{X} = \vec{x}$ corresponds to a conjunction of those primitive events $\bigwedge X_i = x_i$ for $1 \leq i \leq n$.

Let us now evaluate a conditional with disjunctive antecedent in the causal model of Sartorio's Switch. We check whether or not $\langle M, t = 1\rangle \models [F = 1 \vee R = 1]FB = 1$. Let ϕ_1, ϕ_2, ϕ_3 express the disjunctive situations of $F = 1 \vee R = 1$. According to clause (e), we need to check whether $v(D = 1) = 1$ in each $\langle M_{\phi_i}, t = 1\rangle$ for $i = 3$. Figure 2 depicts the causal network of the submodel M_{ϕ_1} for the disjunctive situation ϕ_1. M_{ϕ_2} and M_{ϕ_3} look the same for $\phi_2 = (F = 1) \wedge (R = 0)$ and $\phi_3 = (F = 0) \wedge (R = 1)$.

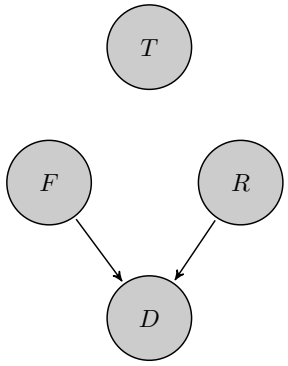


Figure 2: The causal network of M_{ϕ_1} for $\phi_1 = (F = 1) \wedge (R = 1)$.

As $D = max\{F, R\}$ remains unchanged in each $\langle M_{\phi_i}, t = 1\rangle$, we obtain for the three submodels:

(i) $\langle M_{\phi_1}, t = 1\rangle \models D = 1$

(ii) $\langle M_{\phi_2}, t = 1\rangle \models D = 1$

(iii) $\langle M_{\phi_3}, t = 1\rangle \models D = 1$

Hence, $v(D = 1) = 1$ in each $\langle M_{\phi_i}, t = 1\rangle$, and thus the model M satisfies the conditional $[F = 1 \vee R = 1]D = 1$ in context $t = 1$.

3.2 A Refinement of Halpern and Pearl's Definition of Actual Causation

Now that we can evaluate disjunctive antecedents in extended causal models, we propose a refinement or amendment of Halpern and Pearl [2]'s definition of actual causation.

Let $\psi_{\vee_i} = (\vec{X}_i = \vec{x}_i)$ denote the i-th disjunct of the arbitrary Boolean combination ψ of finitely many primitive events.

Definition 7. Actual Causation Refined
ψ is an actual cause of ϕ in $\langle M, \vec{u}\rangle$ iff the following three conditions hold:

AC1R. $\langle M, \vec{u}\rangle \models (\bigwedge \psi_{\vee_i}) \wedge \phi$ for all i.

AC2R. There exists a partition $\langle \vec{Z}, \vec{W}\rangle$ of $\mathcal{V}$ with $\vec{X}_i \subseteq \vec{Z}$ and some setting $\langle \vec{x'}_i, \vec{w'}\rangle$ of the variables in $\langle \vec{X}_i, \vec{W}\rangle$ such that if $\langle M, \vec{u}\rangle \models Z = z^*$ for all $Z \in \vec{Z}$, then both of the following conditions hold:

(a) $\langle M, \vec{u}\rangle \models [\bigwedge \vec{X}_i = \vec{x'}_i, \vec{W} = \vec{w'}]\neg\phi$ for all i.

(b) $\langle M, \vec{u}\rangle \models [\bigvee \vec{X}_i = \vec{x}_i, \vec{W'} = \vec{w'}, \vec{Z'} = \vec{z^*}]\phi$ for all subsets $\vec{W'}$ of $\vec{W}$ and all subsets $\vec{Z'}$ of $\vec{Z}$, and for all i.

AC3R. ψ is minimal; no subsets of the disjuncts $\psi_{\vee_i} = (\vec{X}_i = \vec{x}_i)$ satisfy conditions AC1R and AC2R, and no disjunction of the form $\bigvee(\vec{X}_i = \vec{x}_i) \vee \vec{Y} = \vec{y}$ with $\vec{Y} \subseteq \vec{Z}$, $\vec{Y} \cap \vec{X}_i = \emptyset$ (for all i) and $\vec{Y} \neq \phi$ satisfies AC1R and AC2R.

AC1R requires that each disjunct of the actual cause ψ and its effect ϕ are true in the actual contextualized model. Note that this is equivalent to the big *conjunction* of all disjuncts $\phi_{\vee_i}$ and the effect ϕ being true in the actual contextualized model. (The need for the big conjunction directly follows from Sartorio's first condition necessary for disjunctive causes.)

AC2R requires that each disjunct $\psi_{\vee_i} = (\vec{X}_i = \vec{x}_i)$ of ψ satisfies AC2. That is: (a) setting $\vec{X}_i = \vec{x}_i$ (for any i) changes ϕ to $\neg\phi$, if the variables $\vec{W}$ not on the active causal path(s) take on certain values; (b) guarantees that the disjunction $\bigvee(\vec{X}_i = \vec{x}_i)$ alone is sufficient to change ϕ to $\neg\phi$. Note that AC2R(b) is quite demanding: setting $\bigvee(\vec{X}_i = \vec{x}_i)$ results in a submodel for each disjunctive situation of ψ, and under all of these submodels ϕ is satisfied.

AC3R extends the motivation behind AC3, which is to "prune inessential elements" from the actual causes. The extension demands that if we have another actually occurring disjunct that would alone be sufficient to result in the effect, we need to add it to the disjunctive cause. Correspondingly, we obtain that a formula of the form $(\vec{X} = \vec{x}) \wedge (\vec{Y} = \vec{y})$ for $\vec{X} \cap \vec{Y} = \emptyset$ is more specific and less minimal than $\vec{X} = \vec{x}$, which is in turn more specific and less minimal than $(\vec{X} = \vec{x}) \vee (\vec{Y} = \vec{y})$. Assume this disjunction is an actual cause of some effect. Then the disjunction strips the "overspecific detail" which specific disjunct is causally efficacious (both are!) from the actual cause.

We show now that in $\langle M, t = 1\rangle$ the disjunction $F = 1 \vee R = 1$ is an actual cause of $D = 1$ according to our refined definition. AC1R is satisfied, as $\langle M, t = 1\rangle \models (F = 1 \wedge R = 1) \wedge \phi$. AC2R is satisfied as well. To see this, let $\vec{Z} = \{F, R, D\}$, and thus $\vec{W} = \emptyset$. Clearly, $F, R \subseteq \vec{Z}$. But then (a) $\langle M, t = 1\rangle \models [F = 0 \wedge R = 0]D = 0$. Furthermore, (b) $\langle M, t = 1\rangle \models [F = 1 \vee R = 1]D = 1$, as we have seen in the previous section. Finally, AC3R is satisfied: no subsets of the disjuncts $F = 1$ and $R = 1$ satisfy AC1R and AC2R; there exists no further disjunct satisfying AC1R and AC2R, as $\vec{Z} \setminus \{F, R\} = \{D\}$ and D is the effect.

According to our refined definition, $F = 1$ does not qualify any more as an actual cause of $D = 1$. (The same holds mutatis mutandis for $R = 1$.) The reason is AC3R: $F = 1$ is not minimal. Why? Because there is a disjunction $F = 1 \vee R = 1$ with $R \subseteq \vec{Z}$, $R \cap F = \emptyset$ and $(R = 1) \neq (D = 1)$ satisfying AC1R and AC2R. Hence, $F = 1$ is "inessential" for $D = 1$ in the sense that it is not required for $D = 1$ to obtain, as the actual event $R = 1$ alone would also be sufficient for $D = 1$ to obtain.

4 Conclusion

We generalized Halpern and Pearl [2]'s causal model semantics to allow disjunctive causes of the type found in Sartorio [4]'s Switch. These disjunctive causes have an actual part, i. e. both disjuncts actually occur, and a counterfactual part, i. e. each disjunct would be sufficient for the effect to occur. Based on the causal model semantics extended by disjunctive antecedents à la Sartorio, we refined Halpern and Pearl's definition of actual causation. Halpern and Pearl's original definition qualifies Flipper's flipping the switch as an actual cause of the captivated person's death and does not allow for disjunctive causes. In contrast, our refined definition disqualifies the individual disjuncts as actual causes but makes Sartorio's disjunction "at least one of Flipper flips the switch and Reconnecter reconnects" an actual cause of the person's death. Our refined definition, therefore, implements Sartorio [4, p. 530]'s observation that "there is no general motivation for believing that, when (if) a disjunctive fact is a cause, at least one of its disjuncts must also be a cause."

References

[1] Eiter, T. and Lukasiewicz, T. (2002). Complexity results for structure-based causality. *Artificial Intelligence* **142**(1): 53 – 89. doi:http://dx.doi.org/10.1016/S0004-3702(02)00271-0. URL `http://www.sciencedirect.com/science/article/pii/S0004370202002710`.

[2] Halpern, J. Y. and Pearl, J. (2005). Causes and Explanations: A Structural-Model Approach. Part I: Causes. *British Journal for the Philosophy of Science* **56**(4): 843–887.

[3] Lewis, D. (1973). Causation. *Journal of Philosophy* **70**(17): 556–567.

[4] Sartorio, C. (2006). Disjunctive Causes. *Journal of Philosophy* **103**(10): 521–538.

From Programs to Causal Models*

Thomas F. Icard

Stanford University, Stanford, CA, USA
icard@stanford.edu

Abstract

The purpose of the present contribution is to explore the consequences of building causal models out of programs, and to argue that doing so has advantages for the semantics of subjunctive conditionals and of causal language. We establish basic results about expressivity and give examples to show both the power of the framework and the ways in which it differs from more familiar causal frameworks such as structural equation models.

1 Motivation

The idea that we represent causal relationships with internal "simulation" models has a long and distinguished history, arguably going back to Hume. Perhaps the most prominent contemporary formalization of this idea involves causal Bayesian networks, which define *generative models* over some fixed set of random variables. While Bayes nets are useful for many purposes, some authors have advocated for a more general formalism, known as structural equation models (SEMs), which explicitly encode functional dependencies among variables and relegate all randomness to so called exogenous variables [12]. Unencumbered by the demand for a non-circular account of causal claims, a number of recent researchers in philosophy, linguistics, and psychology have proposed analyzing the semantics of subjunctive conditionals and other ostensibly causal language by appeal to such causal models [16, 9, 14, 17], in place of the once dominant but more abstract "system-of-spheres" models founded on world-similarity-orderings [10].

SEMs come with a number of advantages. By making causal information explicit, they support a precise notion of *intervention*, which grounds hypothetical and counterfactual claims. They can also be applied to a wider array of phenomena than standard Bayes nets, e.g., by allowing certain kinds of cyclic dependencies among variables, which is purportedly important for semantics [5, 14]. Despite these and other attractions, there is a sense in which SEMs depart from the original idea of a simulation model. A prediction in this framework, counterfactual or otherwise, is determined by a solution to a (generally unordered) system of equations, in line with the kinds of models found in physics, economics, and engineering disciplines. But in general structural equations do not simulate; they describe. While this declarative emphasis may be quite appropriate for many purposes, it is desirable to have a similarly expressive framework that retains the procedural character of a simulation model.

A number of authors in artificial intelligence, and more recently in cognitive science, have proposed an idea very much in this vein, to define simulation models using *arbitrary programs* in some rich programming language [13, 11, 4, 3, 1]. Much of the emphasis in this literature is on defining complex probability models with efficient inference procedures. But some authors have also highlighted the fact that these simulation models, just like Bayes nets, may embody causal structure. Despite this important work, a precise analysis of *programs as causal models* has not been given. The purpose of the present contribution is to establish some of the basic definitions and results, and to motivate the idea for semantics of natural language. Rather than

*Thanks to Noah Goodman, Duligur Ibeling, Dan Lassiter, and Krzysztof Mierzewski for helpful discussions.

offer a specific compositional analysis of counterfactuals or causal claims, the aim is to establish the framework with sufficient precision so that any semantic analysis that invokes a notion of "intervention on a simulation model" can be seamlessly accommodated.

As programs themselves have a causal structure, we can use this very structure as a "semi-iconic" causal representation and, as we shall see, as a representation of other non-causal dependence relations as well. The resulting framework provides an attractive setting for a quite general theory of subjunctive conditionals. Highlighted are two especially notable features.

The first is that the framework affords a simple and intelligible way of capturing quantificational and more generally "open-world" reasoning [13, 11], whereby counterfactual suppositions alter which (and even how many) individuals (or other variables) are being considered. The following example is inspired by one from Kaufmann [9, 1164]:

Example 1. Imagine a number of students have shown up to take an exam, and that students typically forget to bring their own pencils. Suppose we say that a student is prepared for the exam just in case either they brought their own pencil, or there are enough pencils for everyone who needs one. (If there are too few, out of fairness no one will be given one.) Upon learning that (1) is true of the situation, does it follows that the counterfactual in (2) is also true?

(1) All of the students are prepared for the exam.

(2) If there had been another five students, they would all be prepared.

This depends on further causal facts: either (1) is true because there is some mechanism in place guaranteeing as many pencils as students, in which case (2) is definitely true; or (1) just happens to be true, in which case (2) could well be false. We would like to model both of these cases, and even the inference about which is more likely—and thus how likely (2) is overall—without having to make specific upfront assumptions about how many students there could be.

A second notable feature is that the move from declarative to procedural emphasis has important logical ramifications, already for propositional logic of counterfactuals.

Example 2. If Alf were ever in trouble, the neighbors Bea and Cam would both like to help. But neither wants to help if the other is already helping. Imagine the following scenario: upon finding out that Alf is in trouble, each looks to see if the other is already there to help. If not, then each begins to prepare to help, eventually making their way to Alf but never stopping again to see if the other is doing the same. If instead, e.g., Cam initially sees Bea already going to help, Cam will not go. One might then argue that (3) and (4) are both intuitively true:

(3) If Alf were in trouble, Bea and Cam would both go to help.

(4) If Alf were in trouble and Bea were going to help, Cam would not go to help.

No existing semantic account of counterfactuals—including both world-ordering models and SEMs—can accommodate this pair of judgments, as $A \boxrightarrow (B \wedge C)$ implies $(A \wedge B) \boxrightarrow C$. The only way to make (3) and (4) both true is to insist that the temporal information be made explicit (evidently unlike typical examples modeled with SEMs [5, 12]). In contrast, by suppressing temporal information in a way that mirrors the surface forms of (3) and (4), it will be easy to find an intuitive simulation making $A \boxrightarrow (B \wedge C)$ true, but $(A \wedge B) \boxrightarrow C$ false.

In what follows we first present the definition of intervention for deterministic programs using Turing machines for concrete illustration, and establish some basic facts about expressivity. We then expand the framework to probabilistic programs so as to handle probabilistic counterfactuals. We consider a number of examples, including Examples 1 and 2, throughout. We also discuss logical and other foundational issues along the way.

2 Intervening on Programs

In thinking of a program as defining a simulation model, we are imagining that there are some variables that initially have some values (the "input") and the program proceeds along, changing these values until it halts, at which point the combination of variable values is construed as the "output" of the simulation. Let us assume programs are Turing machines and that we have a dedicated tape and a fixed interpretation of the tape as a representation of the joint state of infinitely many natural-number-valued variables $\{X_n\}_{n\in\mathbb{N}}$.[1] A *state description* is a set of values $\mathbf{x} = \{x_n\}_{n\in\mathbb{N}}$ for all the variables, only finitely many of which may be non-zero; and a *partial state description* will be any set $\{x_i\}_{i\in I}$, for $I \subseteq \mathbb{N}$. A program can thus be conceived as a (partial) transformation of state descriptions. Let us write $\varphi_{\mathsf{T}}^{\mathbf{x}}(X_i)$ for the value X_i takes on when running machine T on input $\mathbf{x}$, provided T halts (it is undefined otherwise). If we want to consider programs with no (equivalently constantly-0) input, we simply write $\varphi_{\mathsf{T}}(X_i)$.

Of course, the interest in programs as simulation models is not just that they transform inputs to outputs, but that there can be rich dynamics in the course of this transformation. Indeed, a program embodies *counterfactual* information about what *would* happen were we to hold fixed the values of some of the variables throughout the computation.

Definition 1 (Intervention)**.** An *intervention* $\mathcal{I}$ is a computable function that takes (the code of) a program T and produces (code for) a new program $\mathcal{I}(\mathsf{T})$ by selecting a partial state description, $\{x_i\}_{i\in I}$ with $I \subseteq \mathbb{N}$ computable, and holding fixed the values of $\{X_i\}_{i\in I}$ to $\{x_i\}_{i\in I}$ in the computation that T performs. Specifically, $\mathcal{I}$ does the following:

1. Add instructions to the beginning to set the finitely many non-0 variables to their values.
2. Before every instruction α add a routine that checks whether the current cell belongs to a variable X_i with $i \in I$. If $i \notin I$, keep α just as before. If $i \in I$, enter a new state for which there is an instruction just like α, except that the value of the cell is not changed.

Intervening on SEMs involves setting a variable to a given value and then asking what solutions to the equations exist. The intuition here is rather different: intervention on a program involves setting a variable to a given value and letting that manipulation have an effect on the dynamics of the program (i.e., the "simulation"). For a very simple illustration let us return to Example 2. In this example we will help ourselves to "pseudo-code" using `if...then` statements and setting variables to values (writing X `:=` `n` for a number `n`, or X `:=` Y for the current value of variable Y), knowing that we can easily transform all of this into Turing machine code.

Example 3. Let us formalize relevant parts of Example 2 with five binary variables:

B: Bea goes to help	D: Bea intends to help	A: Alf is in trouble
C: Cam goes to help	E: Cam intends to help	

Then consider the following simple program:

```
if A = 1 and C = 0 then D := 1
if A = 1 and B = 0 then E := 1
B := D
C := E
```

[1] Where π is a computable pairing function and $\mathcal{V} = \langle V_n\rangle_{n\in\mathbb{N}}$ is the infinite vector of values on the value tape, let us assume X_i is represented in unary by the infinite sublist $V^{(i)} = \langle V_{\pi(i,1)}V_{\pi(i,2)}V_{\pi(i,3)}\cdots\rangle$. We furthermore assume that programs are written in a normal form so that the value of each X_i is always encoded as a contiguous sequence of 1's followed by the infinite constantly-0 string.

Suppose that in our default initial state all variables are set to 0. It is easy to check that intervening to set $A = 1$ would result in $B = C = 1$. However, if we intervene to set $A = B = 1$, then the program would halt with $C = 0$.

2.1 The Logic of Counterfactual Simulation

One of the main principles in axiomatizations of SEMs is what Pearl calls *composition* [12, 5] (also known as *Cautious Monotonicity* in the literature on non-monotonic logic):

$$(A \boxrightarrow B \wedge A \boxrightarrow C) \Rightarrow (A \wedge B) \boxrightarrow C$$

The declarative character of SEMs, whereby counterfactuals concern finding solutions of equations, establishes this principle as clearly valid: if any solution setting A to 1 would have both B and C set to 1, then any solution that sets A and B to 1 would have C set to 1.

By contrast, on a straightforward construal of what '$\boxrightarrow$' means for programs—intervene to make the antecedent true and see whether the program halts with the consequent true—the composition axiom, while satisfiable, is not valid, as shown by Example 3. At the risk of belaboring the point, the procedural interpretation invokes a very different intuition from the declarative: even though setting $A = 1$ eventually leads to $B = 1$ and $C = 1$, holding $B = 1$ fixed throughout the computation may disrupt the sequence of steps that leads to $C = 1$.

It is possible to give a complete axiomatization of counterfactuals in this setting [6], showing that the logic fails to include several of the validities shared by logics of SEMs and logics interpreted over systems-of-spheres. In a sense, at least concerning the question of which combinations of counterfactual statements can be given a consistent interpretation, the procedural simulation-based perspective can thus be thought of as more general than the declarative SEM approach. For reasons of space, we leave a fuller treatment of these logical issues, and the interpretive questions they raise, for another occasion [6].

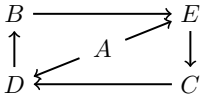
2.2 Defining Causal Graphs

The program in Example 3 clearly reveals an underlying causal structure, which is what supports specific patterns of counterfactuals. Which causal structures can arise from programs? We make this question precise by borrowing a concept from the philosophical literature on causation [18].

Definition 2. Let program T be given. We say that X_i is a *direct cause* of X_j, written $X_i \to X_j$, just in case there are two interventions $\mathcal{I}_1$ and $\mathcal{I}_2$ that hold every variable fixed except for X_j, which differ only in the values assigned to X_i, and for which $\varphi_{\mathcal{I}_1(\mathsf{T})}(X_j) \neq \varphi_{\mathcal{I}_2(\mathsf{T})}(X_j)$.

In other words, $X_i \to X_j$ if X_i directly influences X_j in at least some possible context.

Example 4. Returning to Example 3 it is easy to see that the causal graph defined by this program is as follows:



Note that A does not directly influence B and C, but only via D and E, respectively. Note also that the graph is cyclic, viz. the path $B \to E \to C \to D \to B$.

While in many cases it will be easy to determine the causal graph of a program, the problem in general is unsurprisingly undecidable.

Proposition 1. The problem of determining whether $X_i \to X_j$ is (merely) semi-decidable.

Proof Sketch. That it is semi-decidable is clear: simply dovetail search through all possible pairs of interventions. If there is a pair that results in different values for X_j we will find it.

To see that the problem is not decidable, we reduce it to the problem of determining whether a machine computes a constant function. For any number n consider the Turing machine $\mathsf{T}[n]$ that runs the nth machine T_n on input X_1 and then writes the result of the computation (if it halts) to X_2. We clearly have $X_1 \to X_2$ if and only if T_n does not compute a constant function. As the mapping $n \mapsto \mathsf{T}[n]$ is computable, determining $X_1 \to X_2$ cannot be in general. □

Say that a graph $(\{X_n\}_{n\in\mathbb{N}}, \to)$ is *computably enumerable* (*c.e.*) if the set $\{\langle i, j\rangle : X_i \to X_j\}$ is computably enumerable. It turns out that programs give us all possible c.e. graphs.

Proposition 2. Every c.e. graph is the causal graph for some program.

Proof Sketch. Suppose A is a c.e. set. We describe a program for which $X_i \to X_j$ exactly when $\langle i, j\rangle \in A$. Our program begins by searching to find the first variables with non-zero values (never halting if none is found). Suppose these variables are X_i and X_j with $i < j$.

If $X_i = 1$, then we begin enumerating A until we find the pair $\langle j, i\rangle$ (again, never halting if we never find it). Once found, we write the contents of X_j to X_i and halt. If $X_i > 1$, then we search for the pair $\langle i, j\rangle$ in A. Once found, we write the contents of X_i to X_j and halt.

Clearly, if $\langle n, m\rangle \in A$, then we can find a configuration that witnesses the fact that $X_n \to X_m$. We simply set X_n to either 2 or 3, and X_m to 1 (keeping all others at 0); clearly X_m will depend on X_n no matter whether $n < m$ or $m < n$. If $\langle n, m\rangle \notin A$, then no configuration holding everything but X_n fixed will allow the value of X_m to vary. □

3 Probabilistic Computation and Counterfactuals

To handle causality and counterfactuals in a probabilistic setting we move to stochastic simulations, which we formalize using *probabilistic Turing machines*. In addition to the variable tape encoding $\{X_n\}_{n\in\mathbb{N}}$ we add a *random bit tape* with values $R = \langle R_i\rangle_{i\in\mathbb{N}}$, each bit R_i intuitively representing the result of a fair coin flip.[2] This random source plays a similar role to exogenous variables in SEMs, but in the present context it induces random behavior in our machine: different sequences appearing on the random bit tape may lead to different computations performed by the Turing machine. With T a probabilistic machine, $r \in \{0, 1\}^*$ a finite binary sequence, and $\mathbf{Y}$ a sequence of variables from $\{X_n\}_{n\in\mathbb{N}}$, let us write $\varphi^r_{\mathsf{T}}(\mathbf{Y})$ for the sequence $\mathbf{y}$ of values that variables $\mathbf{Y}$ take on provided T has halted after accessing exactly the random bits of r.[3] Then, as each random bit has probability 2^{-1} and any sequence r has probability $2^{-|r|}$, we can express the probability that machine T halts with values $\mathbf{Y} = \mathbf{y}$ as follows:

$$P_{\mathsf{T}}(\mathbf{Y} = \mathbf{y}) \quad = \quad \sum_{r:\varphi^r_{\mathsf{T}}(\mathbf{Y})=\mathbf{y}} 2^{-|r|}$$

Because machines may have positive probability of not halting at all, the sum over all outputs $\mathbf{y}$ may be less than 1. In this sense P_{T} will be a *semi-measure*. Some authors have suggested limiting attention to machines that almost-surely (with probability 1) halt. It is argued in [7] that this is unnecessarily restrictive, in part because of natural examples like the following.

[2] More formally, the distribution on infinite binary strings is given by the Borel probability space $(\{0, 1\}^\omega, \mathbb{P})$, where $\mathbb{P}$ is the infinite product of Bernoulli(1/2) measures. See, e.g., [3].

[3] Thus, $\varphi^r_{\mathsf{T}}(\mathbf{Y})$ is undefined if T either reads only an initial segment of r or moves beyond r on the random bit tape. Note that given a particular random bit sequence R, the operation of the machine is fully deterministic.

Example 5. Imagine a race between a tortoise and a hare. We have variables $T_0, T_1, T_2, \ldots$ for the position of the tortoise at each time step, and variables $H_0, H_1, H_2, \ldots$ similarly for the hare. Where `Flip(1/4)` is a procedure that returns 1 with probability 1/4 and `Unif(1,7)` returns a number between 1 and 7 uniformly, we might imagine a simulation like this:

T_0 `:= 1;` H_0 `:= 0`
`while (`H_t `<` T_t`)`
    T_{t+1} `:=` T_t `+ 1;` H_{t+1} `:=` H_t
    `if Flip(1/4) then` H_{t+1} `:=` H_t `+ Unif(1,7)`

Whereas this program would almost-surely halt, any small change to the program (e.g., incrementing the tortoise's pace by ϵ) would lead to positive probability of the hare never catching up, even though the two programs may be practically indistinguishable [7].

From a theoretical point of view, we can characterize exactly which semi-measures $P(\mathbf{Y})$ can be defined by a probabilistic Turing machine. We say $P(\mathbf{Y})$ is *enumerable* if for each $\mathbf{y}$ the probability $P(\mathbf{Y} = \mathbf{y})$ can be computably approximated by an increasing sequence of rationals.

Proposition 3 ([7])**.** For every probabilistic Turing machine T, $P_{\mathsf{T}}(\mathbf{Y})$ is an enumerable semi-measure; moreover, every enumerable semi-measure is $P_{\mathsf{T}}(\mathbf{Y})$ for some T.

To capture the causal structure of a probabilistic program we can use the very same definition of intervention (Def. 1). In Example 5, for instance, while $P_{\mathsf{T}}(H_2 \geq T_2) \approx .36$, under an intervention $\mathcal{I}$ that sets H_1 to 1 we would have $P_{\mathcal{I}(\mathsf{T})}(H_2 \geq T_2) \approx .21$. We can also carry over our definition of direct cause (Def. 2) with only slight modification.

Definition 3. Given probabilistic program T, we say $X_i \to X_j$ just in case there are two interventions $\mathcal{I}_1$ and $\mathcal{I}_2$ that hold every variable fixed except for X_j, which differ only in the values assigned to X_i, and for which $P_{\mathcal{I}_1(\mathsf{T})}(X_j) \neq P_{\mathcal{I}_2(\mathsf{T})}(X_j)$.

That is, holding everything but X_j fixed, changing X_i effects a change in probability of X_j.

Example 6. The causal structure of the program in Example 5 consists of two infinite chains:

$$T_0 \longrightarrow T_1 \longrightarrow T_2 \longrightarrow T_3 \longrightarrow \cdots$$

$$H_0 \longrightarrow H_1 \longrightarrow H_2 \longrightarrow H_3 \longrightarrow \cdots$$

As one would expect, the computability and universality results, Props. 1 and 2, apply without change for these probabilistic analogues: we still obtain exactly the c.e. causal graphs.

3.1 Conditioning

Central to the probabilistic setting is the operation of *conditioning* a distribution, which in this context amounts to restricting attention to those runs of the simulation model that eventuate in a particular outcome. Specifically, we can define a (universal) machine COND that takes (codes of) two machines T and F as arguments and (provided F almost-surely halts and returns 1 with positive probability) defines a new simulation model $\mathsf{COND}(\mathsf{T}, \mathsf{F})$ that correctly represents the conditioned semi-measure. For example, if F is a program that checks whether variables $\mathbf{Z}$ would have values $\mathbf{z}$, then $P_{\mathsf{COND}(\mathsf{T},\mathsf{F})}(\mathbf{Y}) = P_{\mathsf{T}}(\mathbf{Y} \mid \mathbf{Z} = \mathbf{z})$, where the latter is defined by the usual ratio formula. This shows that the enumerable semi-measures, or equivalently (by Prop.

3) the machine-definable distributions, are closed under computable conditioning. (See [3] for details on COND and [7] for the general setting of enumerable semi-measures.)

As with other graphical models, conditioning on a "causally upstream" variable is the same as intervening on that variable. For instance, we have $P_{\mathsf{T}}(H_2 \geq T_2 \mid H_1 = 1) = P_{\mathcal{I}(\mathsf{T})}(H_2 \geq T_2)$ in Example 5. The interest comes in combining observations with interventions. Indeed, for Pearl the essence of a counterfactual $A \mathrel{\square\!\!\rightarrow} B$ is captured by a three-step procedure [12, 206]:

1. **Abduction**: Update the model with any relevant observations.
2. **Action**: Modify the model by intervening to make A true.
3. **Prediction**: Use the modified model to compute the probability of B.

Enabling this combination of operations is in fact a major consideration favoring SEMs over Bayes nets, according to Pearl [12, §1.4]. If we like, we can perform the same combination of operations over probabilistic programs.

Example 7. Continuing with Example 5, suppose we observed a run like this:

$$T_0 = 1 \quad H_0 = 0 \qquad T_1 = 2 \quad H_1 = 1 \qquad T_2 = 3 \quad H_2 = 1 \qquad T_3 = 4 \quad H_3 = 4$$

Given the actual trajectory, the hare caught up by time 3 and the simulation terminated. But we could ask, given what happened, if (counter to the facts) the hare had not jumped forward at time 1, would the hare still have caught up by time 3? Where F is a program that verifies the observations above, we first condition T on F to obtain a new program $\mathsf{COND}(\mathsf{T}, \mathsf{F})$. This effectively fixes the first six random choices to ensure that the program (without any interventions) would produce these very observations. However, when we then intervene to set H_1 to 0, running the manipulated program forward results in $H_2 = 0$ and $H_3 = 3$, which means the hare has not caught up and the program would not have halted by time 3.

Given the same observations, and under the same counterfactual supposition, we can also ask what would be the probability of the hare catching up by time 4. If $\mathcal{I}$ is the intervention setting $H_1 = 0$, this is given by $P_{\mathcal{I}(\mathsf{COND}(\mathsf{T},\mathsf{F}))}(H_4 \geq T_4)$, which happens to be 3/14.

3.2 A Note on D-separation and Conditional Independence

Much of the interest in graphical structures in the literature on probability stems from the possibility of reading off (conditional) independence facts from simple graphical properties. For Bayes nets and SEMs, the critical concept is that of *d-separation*. Roughly speaking, variables $\mathbf{Z}$ d-separate $\mathbf{X}$ from $\mathbf{Y}$ if every possible path of information flow from $\mathbf{Y}$ to $\mathbf{X}$ is blocked by some variable in $\mathbf{Z}$.[4] This guarantees that, conditional on $\mathbf{Z}$, $\mathbf{X}$ is independent of $\mathbf{Y}$; that is, $P(\mathbf{X} \mid \mathbf{Z}) = P(\mathbf{X} \mid \mathbf{Y}, \mathbf{Z})$ (see, e.g., [12]).

How does this look for causal graphs defined by programs? Fixing program T, let us say that X_i depends on the nth random bit, written $R_n \to X_i$, just in case there is an intervention $\mathcal{I}$ and sequences r_1 and r_2 that differ only at the nth place, such that $\varphi^{r_1}_{\mathcal{I}(\mathsf{T})}(X_i) \neq \varphi^{r_2}_{\mathcal{I}(\mathsf{T})}(X_i)$. Evidently, if $R_n \to X_i$ and $R_n \to X_j$ this may induce a dependence between them even when X_i and X_j are d-separated in the context of graph $(\{X_n\}_{n\in\mathbb{N}}, \to)$.

If we want d-separation to guarantee (conditional) independence, we have two obvious choices. One is to include $\{R_n\}_{n\in\mathbb{N}}$ as variables in the graph alongside $\{X_n\}_{n\in\mathbb{N}}$ and expand

[4] Specifically, for every path from $\mathbf{Y}$ to $\mathbf{X}$ there must be three variables U, V, W along the path such that either (1) $U \to V \to W$ or $U \leftarrow V \to W$, and $V \in \mathbf{Z}$, or (2) $U \to V \leftarrow W$ and no descendent of V is in $\mathbf{Z}$.

the edge relation $\rightarrow$ accordingly. The other is to insist that we only write programs in such a way that no random bit is a direct cause of two different variables. A similar stipulation is often made in the context of SEMs [12, §1.4]. It is clear that Prop. 2 would not be affected by such a requirement (since that did not require use of the random source at all), but it is perhaps an interesting question whether the universality result in Prop. 3 would still hold. At any rate, either of these stipulations allows for essentially the same argument as for Bayes nets or for SEMs to show conditional independence.

4 Open-World Reasoning

A hallmark of ordinary reasoning in natural language is our ability to deal with situations at a level of abstraction that does not depend on knowing which, or how many, individuals pertain to a given situation. There is no claim that this kind of reasoning is impossible to formalize in other frameworks; the point to emphasize is rather that this kind of reasoning is very natural for simulations built using familiar programming tools such as recursion [13, 11, 4]. In this section we return to consider how one might model the situation described in Example 1.

Suppose we have the following variables, with their intended meanings:

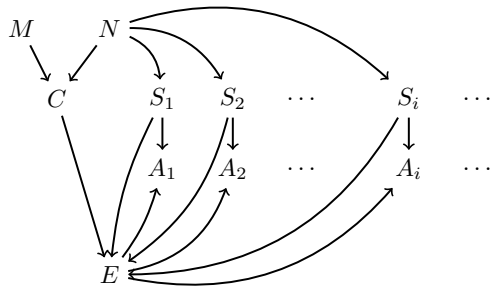
N: number of students	$S_1, S_2, \ldots S_i \ldots$: student i brought their own pencil
M: a mechanism is in place to guarantee the same number of pencils as students	$A_1, A_2, \ldots A_i \ldots$: student i is prepared
C: number of extra pencils	E: There are enough pencils

Let us assume M and E take on values 0 (false) and 1 (true), while variables S_i and A_i take on three values: 0 ("undefined"), 1 (false), and 2 (true). Intuitively A_i should be defined exactly when S_i is, and that should happen only when there actually is an ith student.

In the following program T we assume that four routines for generating numbers randomly are given: $\mathcal{D}_M$, $\mathcal{D}_N$, $\mathcal{D}_C$, and $\mathcal{D}_S$. These can be thought of as defining the "prior" generating procedures for the relevant variables; the precise details will not matter for this example.

M := $\mathcal{D}_M$; N := $\mathcal{D}_N$
C := if M then N else $\mathcal{D}_C$
for i from 1 to N: S_i := $\mathcal{D}_S$
E := $C \geq |\{i : S_i = 1\}|$
for i from 1 to N: A_i := max(S_i,E+1)

The causal graph for T would then look like this:



What does it mean to say that all of the students are prepared (as in (1) from Example 1)? It simply means that none of the variables A_i have value 1, a statement that can be easily checked. Suppose that we know $M = 1$—that there are definitely enough pencils for everyone—and we learn that everyone is prepared. (In fact, the latter is already guaranteed by $M = 1$.) Conditioning our model with a program F_1 that represents these two observations, it is easily seen that the counterfactual in (2), "If there had been another five students, they would all be prepared," has probability 1 according to the program $\mathsf{COND}(\mathsf{T}, \mathsf{F}_1)$.

Suppose on the other hand that we knew $M = 0$. Suppose also that $\mathcal{D}_C$ typically produces small numbers relative to $\mathcal{D}_N$—meaning that there are normally many more students than extra pencils—and that $\mathcal{D}_S$ is such that students almost never bring their own pencils. In such a case learning (1) would be quite surprising. We would moreover expect that there just happened to be enough pencils, but had there been any more students there would not have been enough. Thus, given the conditioned program $\mathsf{COND}(\mathsf{T}, \mathsf{F}_2)$ and where $\mathcal{I}$ is the intervention setting N to $N + 5$, the statement, "All of the students are prepared," has low probability according to $\mathcal{I}\big(\mathsf{COND}(\mathsf{T}, \mathsf{F}_2)\big)$. That is, according to $\mathsf{COND}(\mathsf{T}, \mathsf{F}_2)$, the counterfactual (2) has low probability.

What if we did not know anything about M at all, but merely learned (1). Let F_3 represent observation of (1). As (1) is fully expected when M is true, but quite surprising when M is false, ordinary Bayesian reasoning shows that $\mathsf{COND}(\mathsf{T}, \mathsf{F}_3)$ will now assign M higher probability: in effect, M will now be drawn from a distribution $\mathcal{D}'_M$ that puts more weight on 1 than $\mathcal{D}_M$. The probability of (2) will be intermediate between 1 (the prediction if $M = 1$) and the prediction when $M = 0$, with the precise weighting depending on $\mathcal{D}'_M$, as it intuitively should be.

Two features of this example should be highlighted. The first is, once again, reasoning about this situation does not require making any fixed assumption about which individuals are present. The second is that, though we depict all dependence relations with the same arrow $\rightarrow$, the program T embodies some rather different relationships among variables. For instance, M does not "cause" C in any ordinary sense, but rather modulates whether C depends on N. Similarly, N determines S_i just in the sense that it determines whether there is a student i at all; intuitively, N modulates whether the variable S_i is a relevant part of the current simulation. (Note that we could have made N depend on all of the S_i's.) Philosophers have recognized deep similarities between causal and other kinds of dependence [15]. Because $\rightarrow$ is defined simply by reference to the causal structure of the program, we have blurred all such distinctions.

5 Conclusion

The objective of this paper has been to clarify some of what it might mean to use programs to define causal models, and to ground causal and counterfactual language. In many cases—evidently including most examples considered in recent work in semantics that invoke causal models—the difference between the present framework and more familiar frameworks, such as suitably general classes of SEMs, will not matter. Nonetheless, we have highlighted some important differences, some of which surface already at the level of basic logical validities.

There is certainly no claim that the present framework captures *causality* better than other frameworks. For understanding causal explanation in science, for example, SEMs may often be more useful (see, e.g., [18]). At the same time, for understanding ordinary causal judgments—which have their own distinctive character, cf. [2]—one might argue that the role of simulation is fundamental [4, 3, 1]. Insofar as this is true, we would expect it to be reflected in how people speak about causation as well. At least two points are worth mentioning on this theme.

First, as just mentioned, the framework blurs the distinction between causal and non-causal counterfactuals. While the empirical literature clearly shows that people discriminate causation

from statistical association, it is less obvious that there is any fundamental *cognitive* distinction between causal and, say, logical dependence (or, relatedly, explanation). Within the framework explored here, causal counterfactuals ("If the vase had dropped, it would have broken") can be treated in the very same way as non-causal counterfactuals ("If the vase had been turquoise, then it would have been blue"), which is especially convenient when we need to treat counterfactuals that combine causal and other kinds of dependence (such as sentence (2) from Example 1).

Second, the general framework fits in nicely with a view according to which people select and evaluate counterfactuals stochastically, over richly structured representations, in such a way that the relevant simulation probabilities reflect psychological biases (availability, anchoring, etc.) that can have little to do with "objective" statistics of a situation. This allows incorporating well known psychological effects right into the analysis of conditionals and causal language. As an example, it is often observed that moral considerations affect the way people construct counterfactual scenarios, and, presumably associated with this, their judgments of actual cause ("what caused what"). For instance, the very same act can be judged as more or less causal depending on how people judge its moral status. Recently proposed explanations of these and related phenomena fit very harmoniously with the framework explored here [8].

References

[1] Nick Chater and Mike Oaksford. Programs as causal models: Speculations on mental programs and mental representation. *Cognitive Science*, 37(6):1171–1191, 2013.

[2] David Danks. *Unifying the Mind: Cognitive Representations as Graphical Models*. MIT, 2014.

[3] Cameron E. Freer, Daniel M. Roy, and Joshua B. Tenenbaum. Towards common-sense reasoning via conditional simulation: Legacies of Turing in artificial intelligence. In R. Downey, editor, *Turing's Legacy*. ASL Lecture Notes in Logic, 2012.

[4] Noah D. Goodman, Joshua B. Tenenbaum, and Tobias Gerstenberg. Concepts in a probabilistic language of thought. In Eric Margolis and Stephan Laurence, editors, *The Conceptual Mind: New Directions in the Study of Concepts*. MIT Press, 2015.

[5] Joseph Y. Halpern. Axiomatizing causal reasoning. *Journal of AI Research*, 12:317–337, 2000.

[6] Duligur Ibeling and Thomas Icard. On the logic of counterfactual simulation. Manuscript, 2017.

[7] Thomas Icard. Beyond almost-sure termination. In *Proc. 39th CogSci*, 2017.

[8] Thomas Icard, Jonathan Kominsky, and Joshua Knobe. Normality and actual causal strength. *Cognition*, 161:80–93, 2017.

[9] Stefan Kaufmann. Causal premise semantics. *Cognitive Science*, 37(6):1136–1170, 2013.

[10] David Lewis. *Counterfactuals*. Harvard University Press, 1973.

[11] Brian Milch, Bhaskara Marthi, Stuart Russell, David Sontag, Daniel L. Ong, and Andrey Kolobov. BLOG: Probabilistic models with unknown objects. In *Proc. 19th IJCAI*, pages 1352–1359, 2005.

[12] Judea Pearl. *Causality*. Cambridge University Press, 2009.

[13] Avi Pfeffer and Daphne Koller. Semantics and inference for recursive probability models. In *Proc. 7th National Conference on Artificial Intelligence (AAAI-00)*, pages 538–544, 2000.

[14] Paolo Santorio. Interventions in premise semantics. *Philosophers' Imprint*, 2018.

[15] Jonathan Schaffer. Grounding in the image of causation. *Phil. Studies*, 173(1):49–100, 2016.

[16] Katrin Schulz. "If you'd wiggled A, then B would've changed": Causality and counterfactual conditionals. *Synthese*, 179(2):239–251, 2011.

[17] Steven Sloman, Aron K. Barbey, and Jared M. Hotaling. A causal model theory of the meaning of *cause*, *enable*, and *prevent*. *Cognitive Science*, 33(1):21–50, 2009.

[18] James Woodward. *Making Things Happen: A Theory of Causal Explanation*. OUP, 2003.

Complex antecedents and probabilities in causal counterfactuals*

Daniel Lassiter[1]

Stanford University, Stanford, California, USA

Abstract

Ciardelli, Zhang, & Champollion [5] point out an empirical problem for theories of counterfactuals based on maximal similarity or minimal revision involving negated conjunctions in the antecedent. They also show that disjunctions and negated conjunctions behave differently in counterfactual antecedents, and propose an attractive solution that combines Inquisitive Semantics [4] with a theory of counterfactuals based on interventions on causal models [20]. This paper describes several incorrect empirical predictions of the resulting account, which point to a very general issue for interventionist theories: frequently the antecedent does not give us enough information to choose a unique intervention. The problem applies also to indefinites and to the negation of any non-binary variable. I argue that, when there are multiple ways of instantiating a counterfactual antecedent, we prefer scenarios that are more likely given general probabilistic causal knowledge. A theory is proposed which implements this idea while preserving [5]'s key contributions.

> If I were not a physicist, I would probably be a musician. I often think in music. I live my daydreams in music. I see my life in terms of music.
> — Albert Einstein

1 Introduction

Interventionist theories of counterfactuals have been prominent in recent years in computer science, philosophy of science, statistics, psychology, and many other fields. While many have contributed to this enterprise, Pearl's *Causality* [20] is the most influential document by far. Pearl proposes that counterfactual reasoning proceeds by mutating a model of the causal structure of the world to render the antecedent true, and then considering what follows by causal laws. Semanticists and philosophers of language have begun to explore this approach as well (e.g., [5, 9, 10, 11, 21, 22]). While very attractive, the interventionist semantics is not as well-developed for linguistic purposes as theories based on similarity [15] or premise sets [12, 24].

Most critical, perhaps, is the need to deal seriously with the problem of complex antecedents. If Einstein had said *If I were a musician ...*, the necessary intervention would be fairly clear: we mutate the causal model to make Einstein a musician, and observe what the effects of this change are. But the interventionist semantics does not tell us what to do with his daydream *If I were not a physicist ...*. The problem is just that there are too many alternative professions. When we mutate the causal model so that Einstein is not a physicist, should we make him a barber? an electrician? a musician? unemployed? How can we choose among this bewildering variety of options? Worse, what are we to make of the *probably* in the consequent—if we want Einstein's claim to come out true, do we somehow intervene *non-determistically*, making him

*Many thanks to Lucas Champollion and Thomas Icard for numerous conversations which helped me to get clearer on these issues. Thanks also to Ivano Ciardelli and audiences at UC Davis Language Sciences and the New York Philosophy of Language Workshop.

a musician *most* of the time but sometimes something else? Pearl's semantics is silent on these questions. Failure to treat complex antecedents imposes severe limits on the linguistic generality of the inteventionist approach. These restrictions may well be unproblematic for some modeling purposes, but they are not acceptable if the interventionist semantics is to be linguistically respectable—and to vie with accounts based on similarity or premise sets.

In a recent paper Ciardelli, Zhang, & Champollion [5]—henceforth "CZC"—make a number of important contributions to this problem. First, they show experimentally that negated conjunctions in the antecedent do not behave as expected under maximal similarity/minimal revision theories (see also [2]). Second, they demonstrate the value of the interventionist semantics by providing a natural extension of Pearl's semantics to complex Boolean antecedents that makes better predictions for the negated-conjunction examples. Third, they show that disjunctions and classically equivalent negated conjunctions behave differently, thus motivating the use of Inquisitive Semantics, in which only disjunctions are inquisitive.

However, the proposal also has certain limitations. It makes incorrect predictions about certain counterfactuals with disjunctive and negated antecedents, including some negated conjunctions. In addition, the obvious extension of CZC's propositional semantics to negated indefinites and universals makes strikingly incorrect predictions in some cases.

I will suggest a fix that maintains the core of CZC's proposal, but makes use of a more elaborate way of choosing interventions on the basis of the material in the antecedent. Instead of requiring (in effect) that every way of intervening to render the antecedent true also makes the consequent true, we choose interventions probabilistically, by reasoning about how the antecedent could have come about given the information encoded in the causal model.

2 Non-classical disjunction and causal counterfactuals

CZC experimentally demonstrate a failure of intersubstitutability of classically equivalent propositions in counterfactual antecedents. Consider the scenario **Two Switches**: binary switches A and B are configured so that a light is on (L) iff both are in the same position ($A \wedge B$ or $\neg A \wedge \neg B$). Right now both are up, and the light is on ($A \wedge B \wedge L$).

(1) a. If switch A or switch B were not up, the light would be off. $[\neg A \vee \neg B > \neg L]$
b. If switch A and switch B were not both up, the light would be off. $[\neg(A \wedge B) > \neg L]$

Most experimental participants who saw (1a) judged it true, but most who saw (1b) judged it false or indeterminate. This is despite the fact that (1a) and (1b) are classically equivalent.

CZC account for these examples in two steps. First, they adopt Inquisitive Semantics [4], in which disjunctions are inquisitive but negated conjunctions are not. As [3] describes in detail, Inquisitive Semantics predicts that the default reading for conditionals with disjunctive antecedents will validate "Simplification of Disjunctive Antecedents" (SDA) ([16, 19], etc.; see [1] for a similar Alternative Semantics theory). SDA is the entailment from *If ϕ or ψ, then χ* to *If ϕ, then χ, and if ψ, then χ*. This is enough to account for the preference for "true" in (1a).

Since negated conjunctions are not inquisitive in Inquisitive Semantics, we do not expect SDA in (1b). However, the example is still problematic: if theories of counterfactuals based on minimal revision or maximal similarity were to simply go Inquisitive, they would continue to make incorrect predictions for (1b). The fact that most participants judged (1b) false or indeterminate indicates that, when reasoning about the counterfactual supposition that A and B are not both up $[\neg(A \wedge B)]$, they consider the possibility that the reason that they are not both up is that both are down $[\neg A \wedge \neg B]$. Since this configuration would result in the light still being on, participants do not endorse (1b) unreservedly. However, $\neg A \wedge \neg B$ does not correspond

to a "minimal" revision of the current scenario, which has $A \wedge B$—at least, not in any intuitive sense of "minimal". There are two more minimal revisions: either turn A off and leave B on, or turn B off and leave A on. Both of these modifications would make the antecedent true while turning the light off. So, a theory based on maximal similarity/minimal revision would seem to predict incorrectly that the possibility of $\neg A \wedge \neg B$ should be ignored, rendering (1b) true.

To deal with (1b), CZC adopt a variant of Pearl's semantics based on interventions on causal models [20]. In their model of **Two Switches** there is one causal law— L is a joint effect of A and B $[L \leftrightarrow (A \leftrightarrow B)]$. There are two contingent facts: A and B. To evaluate a counterfactual, intervene to make the antecedent true and consider what follows by causal laws, pruning facts that contribute to the falsity of the antecedent or depend causally on a fact that does. (This summary is necessarily compressed and informal; see [5] for the technical details.) The counterfactual is true iff the consequent is a logical consequence of the causal laws together with the pruned facts and the antecedent. Put another way, the consequent must be true in all models that are consistent with causal laws, antecedent, and pruned facts.

So, for example, we evaluate (1b) by removing all facts that contribute to the falsity of $\neg(A \wedge B)$—which, in this case, are A and B. As a result, the factual basis is empty. There are three kinds of models consistent with the laws. Some have $A \wedge \neg B$, rendering the consequent $\neg L$ true; some have $\neg A \wedge B$, also rendering $\neg L$ true; and some have $\neg A \wedge \neg B$, rendering $\neg L$ false. Since $\neg L$ fails to be true in all of these models, the counterfactual is not true, as desired.

This result constitutes a substantial improvement on standard theories of counterfactuals (which, absent further elaboration, make the wrong prediction for (1b)) and on Pearl's (which makes no predictions about (1a) or (1b)). However, the requirement that the consequent be true in all models that are consistent with causal laws plus pruned facts turns out to be too strong: there are cases where some of the models seem to matter more than others. I'll present the examples first, and the propose a way to make sense of them in terms of explanatory reasoning.

2.1 First puzzle: Failures of SDA.

The use of intervention makes the type of counter-examples to SDA noted by [18] especially acute for CZC. The basic observation is that, when the disjuncts vary substantially in plausibility, the counterfactual supposition may be biased toward the more plausible disjunct.

(2) If it were raining or snowing in Washington, D.C., it would be raining.

By SDA, this should imply *If it were raining in D.C., it would be snowing*, which is absurd. This has often been taken to refute SDA as a semantic principle, but the issue is subtle. Proponents of SDA have objected that the implication has inappropriate presuppositions [8], and that snow in D.C. is being treated as impossible, so that the implication is vacuously true [23, 25, 26]. However, there are related counter-examples to SDA that can't be dismissed in this way.

(3) If it were raining or snowing in D.C., it's likely, but not certain, that it would be raining.

Both of (3)'s entailments by SDA are unsatisfiable. So, SDA is not generally valid.

(4) a. If it were raining in D.C., it's likely, but not certain, that it would be raining.
b. If it were snowing in D.C., it's likely, but not certain, that it would be raining.

The fact that SDA is not always appropriate is not in itself a problem for CZC: all that is needed is an optional semantic operation that can flatten an inquisitive disjunction into a classical disjunction. When this operation is applied, a disjunctive antecedent is equivalent to a negated conjunction. So (2) should be equivalent to (5), and (3) to (6).

(5) If it weren't both not-raining and not-snowing in D.C., it would be raining.

(6) If it weren't both not-raining and not-snowing in D.C., it's likely, but not certain, that it would be raining.

All of these examples are then interpreted like (1b): we throw out facts that contribute to the falsity of the antecedent—here, ¬**rain** and ¬**snow**—and ask what holds in all consistent models. One consistent model has **rain** ∧ ¬**snow**, rendering **rain** true. Another has ¬**rain** ∧ **snow**, rendering **rain** false. Since **rain** cannot be true in all such models, (2) is necessarily false for CZC even on the interpretation that does not validate SDA. Similarly, (3) will turn out false when the theory is supplemented with a plausible treatment of epistemic operators, which should validate the obvious *If ϕ were the case then it's certain that ϕ would be the case.*

2.2 Second puzzle: Partial retention

In the famous **Firing Squad** scenario, riflemen A and B are ready to execute a prisoner. The colonel gives the order (C), and simultaneously A fires (A) and B fires (B). The prisoner dies (D). The laws implicit in the scenario are $\{C \supset A, C \supset B, (A \vee B) \supset D\}$. Now consider (7):

(7) If A and B hadn't both fired, the prisoner would still have died. $[\neg(A \wedge B) > D]$

For CZC (7) is not true, by the same logic as (1b) in **Two Switches**: one way for the riflemen not to *both* shoot is for them to both refrain from shooting. I find this result unsatisfactory, since I can readily imagine judging (7) true along the following lines: if they had not *both* fired, *one of them* would still have fired, since it's extremely unlikely that both would independently and simultaneously (e.g.) have a rifle malfunction, or decide to risk court-martial by disobeying their colonel. Admittedly, the intuition here is not totally compelling. (I will try to explain why below.) A starker issue is CZC's incorrect prediction that (8)-(9) cannot be true under any circumstances, as long as A and B are independent (given C) and both are possible.

(8) If A and B hadn't both fired, one of them would still have fired. $[\neg(A \wedge B) > (A \vee B)]$

(9) If A and B hadn't both fired, the prisoner would still have died, since they wouldn't *both* have had a rifle malfunction. $[\approx \neg(A \wedge B) > D \wedge \neg(A \wedge B) > (A \vee B)]$

Example (7) may be confounded, for example, by the interpretation of *both* and/or focus. In addition, we might rationalize the fact that A and B did not both fire by backtracking to C, considering the possibility that the colonel did not give the order (so that neither would have fired)—though this strategy would not allow us to make sense of (8)-(9). In any case, the same issues arise with other examples. (10) avoids these confounds and is readily read as being true.

(10) If the colonel had given the order and riflemen A, B, C, D, E, F, G, H, I, and J had not all fired, the prisoner would still have died.

On CZC's account we remove facts contributing to the falsity of the antecedent—A fired, B fired, etc.—and ask if the consequent follows. It does not, since there is a model consistent with the laws where the prisoner survives: the one where none of the riflemen fire.

A related example involving universal quantification makes a similar point. Imagine that we are at a Rolling Stones concert with 90,000 screaming fans. I say to you:

(11) If not all of these people had shown up tonight, there would still be a lot of people here.

This is presumably equivalent to (12), with a negated conjunction in the antecedent:

(12) If it weren't the case that (person 1 showed up and person 2 showed up and ... and person 90,000 showed up), there would still be a lot of people here.

Once we remove all facts contributing to the falsity of the antecedent—*Person i showed up* for $i \in \{1, 2, ..., 90000\}$—the consequent clearly does not follow: what if only 3, or 2, or 1, or 0 people showed up? We need an account of why these scenarios are somehow less prominent in reasoning about the counterfactual than ones that are more similar to the actual situation, where (for example) 80,000 or 89,000 show up.

2.3 Third puzzle: Indefinite and negated non-binary antecedents

I have a beagle. If I had a different kind of dog instead, I'd probably have a schnauzer, though I might have a pug. (I would never have more than one dog at the same time, though.)

This is an unremarkable kind of reasoning, but it is difficult to make sense of within interventionist theories, for two reasons. First, it is unclear what intervention is intended: there are many incompatible ways to instantiate the antecedent *If I had a different kind of dog instead*. Second, it is unclear how to make sense of the *probably . . . might . . .* in the consequent: surely, however we intervene to give me a different kind of dog, it's either a schnauzer or not. (Compare Einstein's "If I were not a physicist, I would probably be a musician" discussed in §1.)

The most comprehensive interventionist treatment of complex antecedents to date (CZC's) does not address indefinite antecedents explicitly. But there is an obvious extension: treat indefinites as disjunctions, which can be inquisitive or not. If the antecedent is inquisitive, it is equivalent (by SDA) to (13a). If it is not it is interpreted roughly as (13b).

(13) a. If I had a bulldog I'd probably have a schnauzer, but I might have a pug; and if I had a schnauzer I'd probably have a schnauzer, but I might have a pug; . . .

b. If my pet were in the set **dog** − **beagle**, I'd probably have a schnauzer, but . . .

(13a) is false, assuming I have at most one dog. For (13b), CZC require that the consequent be true in every way of making the antecedent true—i.e., no matter what kind of non-beagle dog I end up with. This cannot be true either. Even if there were only three dog breeds—beagles, schnauzers, and pugs—(14a) and (14b) could not be true (assuming ≤ 1 dog).

(14) a. If my pet were a schnauzer I'd probably have a schnauzer, but I might have a pug.

b. If my pet were a pug I'd probably have a schnauzer, but I might have a pug.

So, this example should be trivially false whether or not the indefinite antecedent is inquisitive.

This is not just a problem about indefinites. Any negation of a non-binary variable—where there are more than two possible alternatives evoked by a negated antecedent—will be associated with multiple ways to instantiate the antecedent. The most obvious extension to CZC's theory for such cases would be to require that the consequent be true under every value for the antecedent other than the one negated. Unfortunately, this won't work for negated antecedents in general. For instance, it predicts that (15a) should be true only when (15b) is.

(15) a. If I ate less chocolate I'd be thinner.

b. For any possible way of eating less chocolate, if I ate less chocolate in that way I'd be thinner.

(15a) intuitively invokes the most likely, or normal, kinds of scenarios that might play out if I ate less chocolate. (15b) is stronger: it is false unless every possible way of eating less chocolate would make me thinner. This subtle mismatch is apparent in (16), where the (a) sentence is reasonable but the attempted paraphrase in (b) is quite strange.

(16) a. If I ate less chocolate I'd probably be thinner, though I might just drink more to make up for it.

b. For any possible way of eating less chocolate, if I ate less chocolate in that way I'd probably be thinner, though I might just drink more to make up for it.

Somehow, we need to soften the interpretation of counterfactuals to focus on normal situations: requiring truth under *all* ways of intervening to make the antecedent true is too stringent.

3 Proposal: Explanatory intervention choice

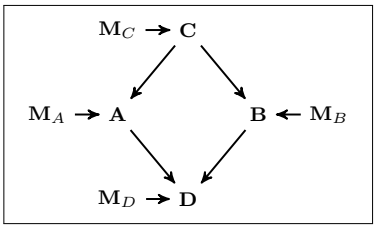
The common feature of our puzzle cases is that different ways of making the antecedent true are not even approximately matched in likelihood. Why is rain the favored instantiation of *rain or snow in D.C.*? Because rain is much more likely than snow in D.C., even though it does snow sometimes. Why, in the concert example, do we prefer to imagine a scenario where the concertgoers do not all show up by letting a smallish number staying home, rather than the entire crowd? The answer has to do with the probabilistic profile of the many, independent decisions that would be involved in the concertgoers all staying home. Since their decisions about whether to come or not are (with localized exceptions) independent, it is plausible enough that a smallish number might have decided to skip the show instead. However, it is very unlikely that a large number would have done so independently. We would have to modify a large number of independent factors to make the consequent false, thus changing the world more radically. This, I suggest, explains why (11) is so plausible.

In both cases, the diagnosis is that our background knowledge about relevant causal forces, and their probabilistic tendencies to produce scenarios compatible with the antecedent, are somehow contributing to the way that we imagine the antecedent being true. The reason that **Two Switches** is different is that the story gives us no insight into how the switches are being set. As a result, we have no basis for concluding that $\neg A \wedge \neg B$ is relatively unlikely, and the scenario where both switches are turned off is given a relatively large weight.

To model the interaction between uncertainty and the interpretation of complex antecedents formally, I will maintain the basic structure behind CZC's theory but switch to using Pearl's [20] Structural Equation Models, which incorporate an explicit representation of probabilistic uncertainty. The information in these models will be used to choose interventions for complex antecedents in a way that emphasizes *explaining* how the intervention could have come about.

To illustrate, consider a model for **Firing squad**. The laws are the same as above, but we write them as structural equations, where "=" represents assignment rather than equality. We also add for each variable V an exogenous source of randomness M_V, with prior probability $P(M_V)$, to represent uncertainty about unmodeled factors that may perturb the otherwise deterministic causal relationships represented in the model. (M is mnemonic for "malfunction".) In this example, the facts $\mathcal{F}$ are $\{C, A, B, D, \neg M_C, \neg M_A, \neg M_B, \neg M_D\}$. The box provides a graphical representation of the causal dependencies represented in the structural equation model.

- $C = \neg M_C$
- $A = C \wedge \neg M_A$
- $B = C \wedge \neg M_B$
- $D = (A \vee B) \wedge \neg M_D$



Each M_X represents a factor that could have perturbed the expected cause/effect relationship. Given that C, A, B and D are true we can infer $\neg M_C, \neg M_A, \neg M_B$, and $\neg M_D$. E.g., A's or B's rifles could have malfunctioned, each with probability p, but they did not.

The proposed procedure for evaluating a counterfactual is as follows. We first prune the facts $\mathcal{F}$ to $\mathcal{F}^*$ as in CZC, removing facts that contribute to the falsity of X or depend on a fact that does. An additional condition is needed to manage the exogenous sources of randomness: for any fact that is pruned, we also throw out the inferred values of any exogenous (M) variables that are immediately relevant to it, resetting their distribution to the prior $P(M)$.[1]

Next we consider all ways of intervening to make the antecedent true. For example, in **Firing Squad** we consider $\{\mathcal{I}_{A\wedge\neg B}, \mathcal{I}_{\neg A\wedge B}, \mathcal{I}_{\neg A\wedge\neg B}\}$, each of which would make *The riflemen do not both fire* true. Conjunctive interventions is treated as sequential intervention.

We then weight the contribution of the various possible interventions to the counterfactual. Here is one method. (There are surely further complexities in the weight function W. The weighted-intervention concept is our main positive contribution, not the precise details of this implementation.) The weight of intervention $\mathcal{I}_X$ is, up to proportionality, $W(\mathcal{I}_X) \propto P(X \mid \mathcal{F}^*)$. We combine the weights of the various possible interventions by normalization. X' ranges over the formulae characterizing the candidate interventions $\mathcal{I}_{X'}$.

$$W(\mathcal{I}_X) = \frac{P(X \mid \mathcal{F}^*)}{\sum_{X'} \; P(X' \mid \mathcal{F}^*)}$$

Normalization means that the weight of an intervention is always relative to other ways of making the antecedent true: a far-fetched possibility might receive high weight nonetheless if the alternatives are even less plausible. Note that the procedure is trivial for simple antecedents: as long as it is causally possible, the unique intervention has weight w that normalizes to $w/w = 1$.

The weight is a measure of the **explanatory value** of the candidate intervention, i.e., the extent to which it does a good job of explaining how the antecedent could have come to be true given the information encoded in the causal model. In essence, the idea is that we prefer ways of making the antecedent true that cohere with the rest of the causal model. This idea is to some extent related to explanatory backtracking (e.g. [6, 17]), but for our purposes we could get away with using backtracking only to *select among* candidate interventions.

Using the weights of the various interventions, we can find the probability of the consequent, given the counterfactual supposition in the antecedent, as the sum of the weights of the candidate interventions that make the consequent true. Some worked-out examples follow. Note that the probabilistic orientation of the proposal gives us an immediate line on the *probably* counterfactuals ((3), (13), etc.) that were troubling for SDA, for the standard interventionist semantics, and for CZC alike: we simply require that the probability assigned to the counterfactual by the method proposed above exceed the relevant threshold (see [13, 14, 27], etc.).

3.1 The Firing Squad and the Stones

In **Firing Squad** we consider *If the riflemen had not both fired, ...* To fix intuitions, let's assume that the colonel will almost certainly give the order: $P(\neg C) = P(M_C) = .01$—while rifle malfunction (willingness to risk court-martial, etc.), is slightly more likely—$P(M_{A/B}) = .1$. $\mathcal{F}$ is $\{A, B, C, D, \neg M_A, \neg M_B, \neg M_C, \neg M_D\}$. All these facts are contribute to the falsity of the antecedent, depend on a fact that does, or contribute randomness to a pruned fact; so, $\mathcal{F}^* = \emptyset$.

[1] This is a first pass. There are other ways that one could manage this issue, and more exploration of complex examples would be needed in order to choose among them.

The candidate interventions are $\mathcal{I}_{A\wedge\neg B}$, $\mathcal{I}_{\neg A\wedge B}$, and $\mathcal{I}_{\neg A\wedge\neg B}$. $W(\mathcal{I}_{\neg A\wedge B}) \propto P(\neg A \wedge B)$, which is just $P(M_A) \times P(\neg M_B)$—$.1 \times .9 = .09$. By analogous reasoning, $W(\mathcal{I}_{A\wedge\neg B}) \propto .09$. For the intervention where neither fires, $W(\mathcal{I}_{\neg A\wedge\neg B})$ is proportional to $P(\neg A \wedge \neg B)$. This is is the sum of the probability that the colonel does not give the order $[P(\neg C) = P(M_C) = .01]$, so that A and B do not fire, and the probability that he does but they fail to fire $[P(\neg M_C) \times P(M_A) \times P(M_B)]$. Using our illustrative values, this means that $W(\mathcal{I}_{\neg A\wedge\neg B}) \propto (.01{+}.99{\times}.1{\times}.1) = .0199$.

Normalizing these values, we find that on these assumptions about prior probabilities $W(\mathcal{I}_{A\wedge\neg B}) = W(\mathcal{I}_{\neg A\wedge B}) \approx .45$, while $W(\mathcal{I}_{\neg A\wedge\neg B}) \approx .1$. Since the prisoner dies in the first two interventions but not the third, the probability of example (7) (*If the riflement had not both fired, the prisoner would still have died*) is equal to $W(\mathcal{I}_{A\wedge\neg B}) + W(\mathcal{I}_{\neg A\wedge B}) \approx .9$. This may help explain the sense that (7) is highly plausible (though not totally compelling) and that its plausibility is related to the striking coincidence—Two simultaneous malfunctions!—that would be required by one of the salient ways keep the prisoner alive.

The model crucially predicts that the probability of (7) is sensitive to $P(C)$, the prior probability that the colonel would give the order. This makes sense: if we had specific knowledge about the colonel—that he must given the order no matter what, or that he is soft-hearted—it may affect our intuitions about the best explanation of *the riflemen do not both fire*. In our model it would have exactly this effect. For instance, if we hold everything the same but make the colonel a softie $[P(\neg C) = .5]$ the best explanation of the riflemen's failure to fire is that the colonel did not give the order. Accordingly, the probability of (7) decreases to about .28.

For the Rolling Stones example (11) (*If not all of these people had shown up, there would still be a lot of people here*), we have to consider a huge number of interventions, each of which removes some particular subset of the 90,000 fans. Suppose that each fan i had probability $P(M_i) = .1$ of deciding not to come, and that each chose independently. Then $W(\mathcal{I}_{\textit{Fan } i \textit{ stays home and the rest come}})$ is .1. If we remove n particular fans, that intervention receives weight $.1^n$. But, for any n, there are $\binom{90000}{n}$ ways for n fans to stay home. The distribution on weights for interventions that remove n fans from the concert is thus:

$$P(\text{If not all of these people had shown up, there would be } n \text{ fewer fans here}) \propto \binom{90000}{n} \times .1^n$$

Figure 1 depicts the distribution on number of fans removed for $P(M_i) = .1$ and $P(M_i) = .9$. Note that the y-axis is in log space: the weight differences are much greater than they appear.

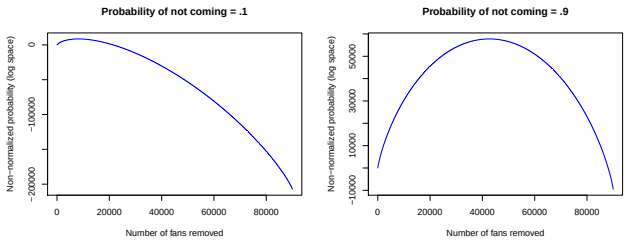


Figure 1: Weight of interventions removing n fans for *If not all of these people had shown up*

The plots show some sensitivity to $P(M_i)$, but the net effect is that there is a strong preference for interventions that remove a relatively small number of fans. The concert remains fairly well-attended ($\approx 40,000$) even if we assume that *all* of the fans were inclined to skip the Stones. The effect is a robust prediction that (11) is highly probable. This seems to be correct.

3.2 Two Switches

The key difference between **Firing Squad** and **Two Switches** is that in the latter case we know nothing about how the switches are set. Here is a simple model, where the positions of switches A and B are controlled by uncorrelated, exogenous causes:

- $A = \neg M_A$
- $B = \neg M_B$
- $L = (A \leftrightarrow B) \wedge \neg M_L$

The facts $\mathcal{F}$ are $\{A, B, L, \neg M_A, \neg M_B, \neg M_L\}$. Knowing nothing of how the switches are set, it is natural to use uninformative priors: $P(M_{A/B}) = .5$. For (1b) $[\neg(A \wedge B) > \neg L]$, all members of $\mathcal{F}$ contribute to the falsity of the antecedent, so $\mathcal{F}^* = \emptyset$. There are three interventions to consider. $W(\mathcal{I}_{A\wedge\neg B}) \propto P(A \wedge \neg B) = P(\neg M_A) \times P(M_B) = .25$. Similarly, $W(\mathcal{I}_{\neg A\wedge B}) \propto P(M_A) \times P(\neg M_B) = .25$, and $W(\mathcal{I}_{\neg A\wedge\neg B}) \propto P(M_A) \times P(M_B) = .25$.

The probability of (1b) is the normalized total weight of interventions that force $\neg L$: $(.25 + .25)/(.25 + .25 + .25) = 2/3$. This middling value may explain why so many participants chose the "Indeterminate" response option in CZC's experiment. By comparison, the disjunction (1a) should (on the SDA reading) have probability 1, and so we expect very high agreement modulo error or noise. This illustrates one way that explanatory reasoning may be able to account for the subtle intuitive differences among **Two Switches**, **Firing Squad**, and the concert example, despite their logical similarity.

3.3 Weather

To model (2)—*If it were raining or snowing in D.C., it would (probably) be raining*—recall that we have to flatten the antecedent to a classical disjunction to avoid inconsistency. Suppose that possible states of weather in D.C. are {**sun, cloud, rain, snow**}, with respective probabilities $.9, .079, .02$, and $.001$]. The only relevant fact, **sun**, is pruned, leaving $\mathcal{F}^*$ empty. Interventions that make **rain** $\vee$ **snow** true are weighted according to prior probabilities: $W(\mathcal{I}_{rain}) = .02$, $W(\mathcal{I}_{snow}) = .001$. (2) is thus true with probability $.02/(.02 + .001) \approx .95$.

4 Conclusion

Interventionist theories of counterfactuals have been hampered by the lack of a treatment of complex antecedents. CZC provide an excellent beginning, but I argued that their requirement of truth in *all* models consistent with the laws and pruned facts—is too strict. I proposed a way of using probabilistic information encoded in Structural Equation Models to weight interventions according to their explanatory value, resulting in a probabilistic interpretation of counterfactuals that maintains the core of CZC's insightful account.

The formal proposal that I have made is resolutely speculative and preliminary. In addition to exploring alternative formalizations, in ongoing work I am testing qualitative predictions regarding, for example, the way that manipulating the causal forces involved in setting the switches in **Two Switches** should influence people's responses, and quantitative predictions about exactly how probabilistic manipulations should do so. Many further questions remain, of course. In addition to the obvious linguistic connections (e.g., counterfactual donkey sentences), there are concerns about the lack of truth-conditions *per se* in the account given here. One possibility is that counterfactuals are thoroughly probabilistic, lacking truth-values (e.g., [7]). Another possibility is that truth could be defined somehow in terms of high probability. I will have to leave these questions for another time.

References

[1] Luis Alonso-Ovalle. Counterfactuals, correlatives, and disjunction. *Linguistics and Philosophy*, 32(2):207–244, 2009.

[2] Lucas Champollion, Ivano Ciardelli, and Linmin Zhang. Breaking de Morgan's law in counterfactual antecedents. In Mary Moroney, Carol-Rose Little, Jacob Collard, and Dan Burgdorf, editors, *26th Semantics and Linguistic Theory Conference (SALT 26)*, pages 304–324, Ithaca, NY, 2016. LSA and CLC Publications.

[3] Ivano Ciardelli. Lifting conditionals to inquisitive semantics. In *Semantics and Linguistic Theory*, volume 26, pages 732–752, 2016.

[4] Ivano Ciardelli, Jeroen Groenendijk, and Floris Roelofsen. Inquisitive semantics: a new notion of meaning. *Language and Linguistics Compass*, 7(9):459–476, 2013.

[5] Ivano Ciardelli, Linmin Zhang, and Lucas Champollion. Two switches in the theory of counterfactuals: A study of truth conditionality and minimal change. *Linguistics and Philosophy*, to appear.

[6] Morteza Dehghani, Rumen Iliev, and Stefan Kaufmann. Causal explanation and fact mutability in counterfactual reasoning. *Mind & Language*, 27(1):55–85, 2012.

[7] Dorothy Edgington. Counterfactuals. In *Proceedings of the Aristotelian Society*, volume 108, pages 1–21, 2008.

[8] Kit Fine. Counterfactuals without possible worlds. *Journal of Philosophy*, 109(3):221–246, 2012.

[9] Eric Hiddleston. A causal theory of counterfactuals. *Noûs*, 39(4):632–657, 2005.

[10] Stefan Kaufmann. *Aspects of the Meaning and Use of Conditionals*. PhD thesis, Stanford, 2001.

[11] Stefan Kaufmann. Causal premise semantics. *Cognitive science*, 37(6):1136–1170, 2013.

[12] Angelika Kratzer. Partition and revision: The semantics of counterfactuals. *Journal of Philosophical Logic*, 10(2):201–216, 1981.

[13] Daniel Lassiter. Gradable epistemic modals, probability, and scale structure. In Nan Li and David Lutz, editors, *Semantics & Linguistic Theory (SALT) 20*, pages 197–215. CLC Publications, 2010.

[14] Daniel Lassiter. *Graded Modality*. Oxford University Press, 2017.

[15] David Lewis. *Counterfactuals*. Harvard University Press, 1973.

[16] Barry Loewer. Counterfactuals with disjunctive antecedents. *The Journal of Philosophy*, 73(16):531–537, 1976.

[17] Christopher G. Lucas and Charles Kemp. An improved probabilistic account of counterfactual reasoning. *Psychological Review*, 122(4):700–734, 2015.

[18] Thomas McKay and Peter Van Inwagen. Counterfactuals with disjunctive antecedents. *Philosophical studies*, 31(5):353–356, 1977.

[19] Donald Nute. Counterfactuals and the similarity of worlds. *Journal of Philosophy*, 72(21):773–778, 1975.

[20] Judea Pearl. *Causality: Models, Reasoning and Inference*. Cambridge University Press, 2000.

[21] Paolo Santorio. Interventions in premise semantics. *Philosophers' Imprint*, 2016.

[22] Katrin Schulz. "If youd wiggled A, then B wouldve changed": Causality and counterfactual conditionals. *Synthese*, 179(2):239–251, 2011.

[23] William B Starr. A uniform theory of conditionals. *Journal of Philosophical Logic*, 43(6):1019–1064, 2014.

[24] Frank Veltman. *Logics for conditionals*. PhD thesis, University of Amsterdam, 1985.

[25] Ken Warmbrōd. Counterfactuals and substitution of equivalent antecedents. *Journal of Philosophical Logic*, 10(2):267–289, 1981.

[26] Malte Willer. Simplifying with free choice. 2017. In press at *Topoi*.

[27] Seth Yalcin. Probability operators. *Philosophy Compass*, 5(11):916–937, 2010.

Lexical and Derivational Meaning in Vector-Based Models of Relativisation

Michael Moortgat[1] and Gijs Wijnholds[2]

[1] Utrecht University, The Netherlands
m.j.moortgat@uu.nl
[2] Queen Mary University of London, United Kingdom
g.j.wijnholds@qmul.ac.uk

Abstract

Sadrzadeh et al (2013) present a compositional distributional analysis of relative clauses in English in terms of the Frobenius algebraic structure of finite dimensional vector spaces. The analysis relies on distinct type assignments and lexical recipes for subject vs object relativisation. The situation for Dutch is different: because of the verb final nature of Dutch, relative clauses are ambiguous between a subject vs object relativisation reading. Using an extended version of Lambek calculus, we present a compositional distributional framework that accounts for this derivational ambiguity, and that allows us to give a single meaning recipe for the relative pronoun reconciling the Frobenius semantics with the demands of Dutch derivational syntax.

1 Introduction

Compositionality, as a structure-preserving mapping from a syntactic source to a target interpretation, is a fundamental design principle both for the set-theoretic models of formal semantics and for syntax-sensitive vector-based accounts of natural language meaning, see [1] for discussion. For typelogical grammar formalisms, to obtain a compositional interpretation, we have to specify how the Syn-Sem homomorphism acts on *types* (basic and complex) and on *proofs* (derivations, again basic (axioms) or compound, obtained by inference steps). There is a tension here between lexical and derivational aspects of meaning: the derivational aspects relate to the composition operations associated with the inference steps that put together phrases out of more elementary parts; the atoms for this composition process are the meanings of the lexical constants associated with the axioms of a derivation.

Relative clause structures form a suitable testbed to study the interaction between these two aspects of meaning, and they have been well-studied in the formal and in the distributional settings. Informally, a restrictive relative clause ('books that Alice read') has an intersective interpretion. In the formal semantics account, this interpretation is obtained by modeling both the head noun ('books') and the relative clause body ('Alice read $\_$') as (characteristic functions of) sets (type $e \to t$); the relative pronoun can then be interpreted as the intersection operation. In distributional accounts such as [2], full noun phrases and simple common nouns are interpreted in the same semantic space, say N, distinct from the sentence space S. In this setting, element-wise multiplication, which preserves non-null context features, is a natural candidate for an intersective interpretation; in the case at hand this means element-wise multiplication of a vector in N interpreting the head noun, with a vector interpretation obtained from the relative clause body. To achieve this effect, [9] rely on the Frobenius algebraic structure of **FVect**, which provides operations for (un)copying, insertion and deletion of vector information. A key feature of their account is that it relies on *structure-specific* solutions of the lexical equation: subject and object relative clauses are obtained from distinct type assignments to the relative

$$\frac{}{1_A : A \longrightarrow A} \qquad \frac{f : A \longrightarrow B \quad g : B \longrightarrow C}{g \circ f : A \longrightarrow C}$$

$$\frac{f : \Diamond A \longrightarrow B}{\nabla f : A \longrightarrow \Box B} \qquad \frac{f : A \otimes B \longrightarrow C}{\triangleright f : A \longrightarrow C/B} \qquad \frac{f : A \otimes B \longrightarrow C}{\triangleleft f : B \longrightarrow A\backslash C}$$

$$\frac{g : A \longrightarrow \Box B}{\nabla^{-1} g : \Diamond A \longrightarrow B} \qquad \frac{g : A \longrightarrow C/B}{\triangleright^{-1} g : A \otimes B \longrightarrow C} \qquad \frac{g : B \longrightarrow A\backslash C}{\triangleleft^{-1} g : A \otimes B \longrightarrow C}$$

$$\alpha^l_\Diamond : \Diamond A \otimes (B \otimes C) \longrightarrow (\Diamond A \otimes B) \otimes C \qquad \alpha^r_\Diamond : (A \otimes B) \otimes \Diamond C \longrightarrow A \otimes (B \otimes \Diamond C)$$
$$\sigma^l_\Diamond : \Diamond A \otimes (B \otimes C) \longrightarrow B \otimes (\Diamond A \otimes C) \qquad \sigma^r_\Diamond : (A \otimes B) \otimes \Diamond C \longrightarrow (A \otimes \Diamond C) \otimes B$$

Figure 1: $\mathbf{NL}_\Diamond$. Residuation rules; extraction postulates.

pronoun (Lambek types $(n\backslash n)/(np\backslash s)$ vs $(n\backslash n)/(s/np)$), associated with distinct instructions for meaning assembly.

For a language like Dutch, such an account is problematic. Dutch subordinate clause order has the SOV pattern Subj–Obj–TV, i.e. a transitive verb is typed as $np\backslash(np\backslash s)$, selecting its arguments uniformly to the left. As a result, example (1)(a) is ambiguous between a subject vs object relativisation interpretation: it can be translated as either (b) or (c). The challenge here is twofold: at the syntactic level, we have to provide a *single* type assignment to the relative pronoun that can withdraw either a subject or an object hypothesis from the relative clause body; at the semantic level, we need a *uniform* meaning recipe for the relative pronoun that will properly interact with the derivational semantics.

a	mannen$_n$ die$_?$ vrouwen$_{np}$ haten$_{np\backslash(np\backslash s)}$	(ambiguous)	
b	men who hate women	(subject rel)	(1)
c	men who(m) women hate	(object rel)	

The paper is structured as follows. In §2, we present an extended version of Lambek calculus, and show how it accounts for the derivational ambiguity of Dutch relative clauses. In §3.1, we define the interpretation homomorphism that associates syntactic derivations with composition operations in a vector-based semantic model. The derivational semantics thus obtained is formulated at the type level, i.e. it abstracts from the contribution of individual lexical items. In §3.2, we bring in the lexical semantics, and show how the Dutch relative pronoun can be given a uniform interpretation that properly interacts with the derivational semantics. The discussion in §4 compares the distributional and formal semantics accounts of relativisation.

2 Syntax

Our syntactic engine is $\mathbf{NL}_\Diamond$ [6]: the extension of Lambek's [3] Syntactic Calculus with an adjoint pair of control modalities $\Diamond, \Box$. The modalities play a role similar to that of the exponentials of linear logic: they allow one to introduce controlled, rather than global, forms of reordering and restructuring. In this paper, we consider the controlled associativity and commutativity postulates of [7]. One pair, $\alpha^l_\Diamond, \sigma^l_\Diamond$, allows a $\Diamond$-marked formula to reposition itself on left branches of a constituent tree; we use it to model the SOV extraction patterns in Dutch. A

symmetric pair $\alpha^r_\diamond, \sigma^r_\diamond$ would capture the non-local extraction dependencies in an SVO language such as English. Lambek [4] has shown how deductions in a syntactic calculus can be viewed as arrows in a category. Figure 1 presents $\mathbf{NL}_\diamond$ in this format.

For parsing, we want a proof search procedure that doesn't rely on cut. Consider the rules in Figure 2, expressing the monotonicity properties of the type-forming operations, and recasting the postulates in rule form. It is routine to show that these are *derived* rules of inference of $\mathbf{NL}_\diamond$. In [8] it is shown that by adding them to the residuation rules of Figure 1, one obtains a system equivalent to a display sequent calculus enjoying cut-elimination. By further restricting to *focused* derivations, proof search is free of spurious ambiguity.

We are ready to return to our example (1)(a). A type assignment $(n\backslash n)/(\Diamond\Box np\backslash s)$ to the relative pronoun 'die' accounts for the derivational ambiguity of the phrase. The derivations agree on the initial steps

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{}{n\longrightarrow n}\quad\dfrac{}{n\longrightarrow n}}{n\backslash n\longrightarrow n\backslash n}\backslash\quad\dfrac{\vdots}{np\otimes(np\backslash(np\backslash s))\longrightarrow\Diamond\Box np\backslash s}}{(n\backslash n)/(\Diamond\Box np\backslash s)\longrightarrow(n\backslash n)/(np\otimes(np\backslash(np\backslash s)))}/}{((n\backslash n)/(\Diamond\Box np\backslash s))\otimes(np\otimes(np\backslash(np\backslash s)))\longrightarrow n\backslash n}\triangleright^{-1}}{n\otimes(((n\backslash n)/(\Diamond\Box np\backslash s))\otimes(np\otimes(np\backslash(np\backslash s))))\longrightarrow n}\triangleleft^{-1} \tag{2}$$

but then diverge in how the relative clause body is derived:

$$\dfrac{\dfrac{\dfrac{}{np\longrightarrow np}\quad\dfrac{\dfrac{\dfrac{\dfrac{}{np\longrightarrow np}}{\Box np\longrightarrow\Box np}\Box}{\Diamond\Box np\longrightarrow np}\nabla^{-1}\quad\dfrac{}{s\longrightarrow s}\backslash}{\boxed{np\backslash s\longrightarrow\Diamond\Box np\backslash s}}}{np\backslash(np\backslash s)\longrightarrow np\backslash(\Diamond\Box np\backslash s)}\backslash}{np\otimes(np\backslash(np\backslash s))\longrightarrow\Diamond\Box np\backslash s}\triangleleft^{-1}$$

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{}{np\longrightarrow np}}{\Box np\longrightarrow\Box np}\Box}{\Diamond\Box np\longrightarrow np}\nabla^{-1}\quad\dfrac{\dfrac{}{np\longrightarrow np}\quad\dfrac{}{s\longrightarrow s}\backslash}{np\backslash s\longrightarrow np\backslash s}\backslash}{\boxed{np\backslash(np\backslash s)\longrightarrow\Diamond\Box np\backslash(np\backslash s)}}}{\Diamond\Box np\otimes(np\backslash(np\backslash s))\longrightarrow np\backslash s}\triangleleft^{-1}}{np\otimes(\Diamond\Box np\otimes(np\backslash(np\backslash s)))\longrightarrow s}\triangleleft^{-1}}{\Diamond\Box np\otimes(np\otimes(np\backslash(np\backslash s)))\longrightarrow s}\widehat{\sigma}^l_\diamond}{np\otimes(np\backslash(np\backslash s))\longrightarrow\Diamond\Box np\backslash s}\triangleleft \tag{3}$$

In the derivation on the left, the $\Diamond\Box np$ hypothesis is linked to the *subject* argument of the verb; in the derivation on the right to the *object* argument, reached via the $\widehat{\sigma}^l_\diamond$ reordering step.

$$\frac{f : A\longrightarrow B}{\Diamond f : \Diamond A\longrightarrow\Diamond B}\qquad\frac{f : A\longrightarrow B}{\Box f : \Box A\longrightarrow\Box B}$$

$$\frac{f : A\longrightarrow B\quad g : C\longrightarrow D}{f\otimes g : A\otimes C\longrightarrow B\otimes D}\qquad\frac{f : A\longrightarrow B\quad g : C\longrightarrow D}{f/g : A/D\longrightarrow B/C}\qquad\frac{f : A\longrightarrow B\quad g : C\longrightarrow D}{f\backslash g : B\backslash C\longrightarrow A\backslash D}$$

$$\frac{f : (\Diamond A\otimes B)\otimes C\longrightarrow D}{\widehat{\alpha}^l_\diamond f : \Diamond A\otimes(B\otimes C)\longrightarrow D}\qquad\frac{f : B\otimes(\Diamond A\otimes C)\longrightarrow D}{\widehat{\sigma}^l_\diamond f : \Diamond A\otimes(B\otimes C)\longrightarrow D}$$

Figure 2: $\mathbf{NL}_\diamond$. Monotonicity; leftward extraction (rule version).

3 From source to target

3.1 Derivational semantics

Compositional distributional models are obtained by defining a homomorphism sending types and derivations of a syntactic source system to their counterparts in a symmetric compact closed category (sCCC); the concrete model for this sCCC then being finite dimensional vector spaces (**FVect**) and (multi)linear maps. Such interpretation homomorphisms have been defined for pregroup grammars, Lambek calculus and CCG in [2, 5]. We here define the interpretation for $\mathbf{NL}_\diamond$, starting out from [10].

Recall first that a *compact closed category* (CCC) is monoidal, i.e. it has an associative $\otimes$ with unit I; and for every object there is a left and a right adjoint satisfying

$$A^l \otimes A \xrightarrow{\epsilon^l} I \xrightarrow{\eta^l} A \otimes A^l \qquad A \otimes A^r \xrightarrow{\epsilon^r} I \xrightarrow{\eta^r} A^r \otimes A$$

In a *symmetric* CCC, the tensor moreover is commutative, and we can write A^* for the collapsed left and right adjoints.

In the concrete instance of **FVect**, the unit I stands for the field $\mathbb{R}$; identity maps, composition and tensor product are defined as usual. Since bases of vector spaces are fixed in concrete models, there is only one natural way of defining a basis for a *dual space*, so that $V^* \cong V$. In concrete models we may collapse the adjoints completely.

The ϵ map takes inner products, whereas the η map (with $\lambda = 1$) introduces an identity tensor as follows:

$$\begin{array}{lll} \epsilon_V : V \otimes V \to \mathbb{R} \quad \text{given by} & \sum_{ij} v_{ij}(\vec{e}_i \otimes \vec{e}_j) & \mapsto \quad \sum_i v_{ii} \\ \eta_V : \mathbb{R} \to V \otimes V \quad \text{given by} & \lambda & \mapsto \quad \sum_i \lambda(\vec{e}_i \otimes \vec{e}_i) \end{array}$$

Interpretation: types At the type level, the interpretation function $\lceil\cdot\rceil$ assigns a vector space to the atomic types of $\mathbf{NL}_\diamond$; for complex types we set $\lceil \Diamond A \rceil = \lceil \Box A \rceil = \lceil A \rceil$, i.e. the syntactic control operators are transparent for the interpretation; the binary type-forming operators are interpreted as

$$\lceil A \otimes B \rceil = \lceil A \rceil \otimes \lceil B \rceil \quad \lceil A/B \rceil = \lceil A \rceil \otimes \lceil B \rceil^* \quad \lceil A \backslash B \rceil = \lceil A \rceil^* \otimes \lceil B \rceil$$

Interpretation: proofs From the linear maps interpreting the premises of the $\mathbf{NL}_\diamond$ inference rules, we want to compute the linear map interpreting the conclusion. Identity and composition are immediate: $\lceil 1_A \rceil = 1_{\lceil A \rceil}$, $\lceil g \circ f \rceil = \lceil g \rceil \circ \lceil f \rceil$. For the residuation inferences, from the map $\lceil f \rceil : \lceil A \rceil \otimes \lceil B \rceil \longrightarrow \lceil C \rceil$ interpreting the premise, we obtain

$$\begin{array}{l} \lceil \triangleright f \rceil = \lceil A \rceil \xrightarrow{1_{\lceil A \rceil} \otimes \eta_{\lceil B \rceil}} \lceil A \rceil \otimes \lceil B \rceil \otimes \lceil B \rceil^* \xrightarrow{\lceil f \rceil \otimes 1_{\lceil B \rceil^*}} \lceil C \rceil \otimes \lceil B \rceil^* \\ \lceil \triangleleft f \rceil = \lceil B \rceil \xrightarrow{\eta_{\lceil A \rceil} \otimes 1_{\lceil B \rceil}} \lceil A \rceil^* \otimes \lceil A \rceil \otimes \lceil B \rceil \xrightarrow{1_{\lceil A \rceil^*} \otimes \lceil f \rceil} \lceil A \rceil^* \otimes \lceil C \rceil \end{array}$$

For the inverses, from maps $\lceil g \rceil : \lceil A \rceil \longrightarrow \lceil C/B \rceil$, $\lceil h \rceil : \lceil B \rceil \longrightarrow \lceil A \backslash C \rceil$ for the premises, we obtain

$$\begin{array}{l} \lceil \triangleright^{-1} g \rceil = \lceil A \rceil \otimes \lceil B \rceil \xrightarrow{\lceil g \rceil \otimes 1_{\lceil B \rceil}} \lceil C \rceil \otimes \lceil B \rceil^* \otimes \lceil B \rceil \xrightarrow{1_{\lceil C \rceil} \otimes \epsilon_{\lceil B \rceil}} \lceil C \rceil \\ \lceil \triangleleft^{-1} h \rceil = \lceil A \rceil \otimes \lceil B \rceil \xrightarrow{1_{\lceil A \rceil} \otimes \lceil h \rceil} \lceil A \rceil \otimes \lceil A \rceil^* \otimes \lceil C \rceil \xrightarrow{\epsilon_{\lceil A \rceil} \otimes 1_{\lceil C \rceil}} \lceil C \rceil \end{array}$$

Monotonicity. The case of parallel composition is immediate: $\lceil f \otimes g \rceil = \lceil f \rceil \otimes \lceil g \rceil$. For the slash cases, from $\lceil f \rceil : \lceil A \rceil \longrightarrow \lceil B \rceil$ and $\lceil g \rceil : \lceil C \rceil \longrightarrow \lceil D \rceil$, we obtain

$$\lceil f/g \rceil = \qquad\qquad \lceil f\backslash g \rceil =$$

$$\begin{array}{c} \lceil A \rceil \otimes \lceil D \rceil^* \\ \downarrow{\scriptstyle \lceil f \rceil \otimes \eta_{\lceil C \rceil} \otimes 1_{\lceil D \rceil^*}} \\ \lceil B \rceil \otimes \lceil C \rceil^* \otimes \lceil C \rceil \otimes \lceil D \rceil^* \\ \downarrow{\scriptstyle 1_{\lceil B \rceil \otimes \lceil C \rceil^*} \otimes \lceil g \rceil \otimes 1_{\lceil D \rceil^*}} \\ \lceil B \rceil \otimes \lceil C \rceil^* \otimes \lceil D \rceil \otimes \lceil D \rceil^* \\ \downarrow{\scriptstyle 1_{\lceil B \rceil \otimes \lceil C \rceil^*} \otimes \epsilon_{\lceil D \rceil}} \\ \lceil B \rceil \otimes \lceil C \rceil^* \end{array} \qquad\qquad \begin{array}{c} \lceil B \rceil^* \otimes \lceil C \rceil \\ \downarrow{\scriptstyle 1_{\lceil B \rceil^*} \otimes \eta_{\lceil A \rceil} \otimes \lceil g \rceil} \\ \lceil B \rceil^* \otimes \lceil A \rceil \otimes \lceil A \rceil^* \otimes \lceil D \rceil \\ \downarrow{\scriptstyle 1_{\lceil B \rceil^*} \otimes \lceil f \rceil \otimes 1_{\lceil A \rceil^* \otimes \lceil D \rceil}} \\ \lceil B \rceil^* \otimes \lceil B \rceil \otimes \lceil A \rceil^* \otimes \lceil D \rceil \\ \downarrow{\scriptstyle \epsilon_{\lceil B \rceil} \otimes 1_{\lceil A \rceil^* \otimes \lceil D \rceil}} \\ \lceil A \rceil^* \otimes \lceil D \rceil \end{array}$$

Interpretation for the extraction structural rules is obtained via the standard associativity and symmetry maps of **FVect**: $\lceil \widehat{\alpha}^l_\diamond f \rceil = f \circ \alpha$ and $\lceil \widehat{\sigma}^l_\diamond f \rceil = f \circ \alpha^{-1} \circ (\sigma \otimes 1_A) \circ \alpha$ and similarly for the rightward extraction rules.

Simplifying the interpretation Whereas the syntactic derivations of $\mathbf{NL}_\diamond$ proceed in cut-free fashion, the interpretation of the inference rules given above introduces detours (sequential composition of maps) that can be removed. We use a generalised notion of Kronecker delta, together with Einstein summation notation, to concisely express the fact that the interpretation of a derivation is fully determined by the identity maps that interpret its axiom leaves, realised as the ϵ or η identity matrices depending on their (co)domain signature.

Recall that vectors and linear maps over the real numbers can be equivalently expressed as (multi-dimensional) arrays of numbers. The essential information one needs to keep track of are the coefficients of the tensor: for a vector $\mathbf{v} \in \mathbb{R}^n$ we write v_i (with i ranging from 1 to n), an $n \times m$ matrix $\mathbf{A}$ is expressed as A_{ij}, an $n \times m \times p$ cube $\mathbf{B}$ as B_{ijk}, with the indices each time ranging over the dimensions. The Einstein summation convention on indices then states that in an expression involving multiple tensors, indices occurring once give rise to a tensor product, whereas indices occurring twice are contracted. Without explicitly writing a tensor product $\otimes$, the tensor product of a vector $\mathbf{a}$ and a matrix $\mathbf{A}$ thus can be written as $a_i A_{jk}$; the inner product between vectors $\mathbf{a}, \mathbf{b}$ is $a_i b_i$. Matrix application $\mathbf{Aa}$ is rendered as $A_{ij} a_j$, i.e. the contraction happens over the second dimension of $\mathbf{A}$ and $\mathbf{a}$. For tensors of arbitrary rank we use uppercase to refer to lists of indices: we write a tensor $\mathbf{T}$ as T_I. Tensor application then becomes $T_{IJ} R_J$, for some tensor $\mathbf{R}$ of lower rank.

The identity matrix is given by the Kronecker delta (left), the identity tensor by its generalisation (right):

$$\delta^i_j = \begin{cases} 1 & i = j \\ 0 & \text{otherwise} \end{cases} \qquad\qquad \delta^I_J = \begin{cases} 1 & I_k = J_k \text{ for all } k \\ 0 & \text{otherwise} \end{cases}$$

The attractive property of the (generalised) Kronecker delta is that it expresses unification of indices: $\delta^i_j a_i = a_j$, which is simply a renaming of the index; the inner product can be computed by $\delta^i_j a_i b_j = a_j b_j$. Left on its own, it is simply an identity matrix/tensor.

With the Kronecker delta, the composition of matrices $\mathbf{B} \circ \mathbf{A}$ is expressible as $\delta^j_k A_{ij} B_{kl}$, which is the same as $A_{ij}B_{jl}$ (or $A_{ik}B_{kl}$). We can show that order of composition is irrelevant:

$$\delta^j_k A_{ij} \delta^l_m B_{kl} C_{mn} = A_{ij} B_{jl} C_{ln} = \delta^l_m \delta^j_k A_{ij} B_{kl} C_{mn}$$

The special cases of tensor product of generalised Kronecker deltas is given by concatenating the index lists:

$$\delta^I_J \otimes \delta^K_L = \delta^{IK}_{JL}$$

expressing the fact that $1_A \otimes 1_B = 1_{A \otimes B}$.

Since the generalised Kronecker delta is able to do renaming, take inner product, and insert an identity tensor, depending on the number of arguments placed behind it, it will represent precisely the $1_A, \epsilon_A, \eta_A$ maps discussed above. In this respect, the interpretation can be simplified and we can label the proof system (with formulas already interpreted) with these generalised Kronecker deltas. The effect of the residuation rules and the structural rules is to only change the (co)domain signature of a Kronecker delta, whereas the rules for axioms and monotonicity also act on the Kronecker delta itself:

$$\frac{}{A_I \xrightarrow{\delta^I_J} A_J}\ 1_A$$

$$\frac{A \xrightarrow{\delta^I_J} B \quad C \xrightarrow{\delta^K_L} D}{A \otimes C \xrightarrow{\delta^{IK}_{JL}} B \otimes D}\ \otimes \qquad \frac{A \xrightarrow{\delta^I_J} B \quad C \xrightarrow{\delta^K_L} D}{A \otimes D \xrightarrow{\delta^{IK}_{JL}} B \otimes C}\ / \qquad \frac{A \xrightarrow{\delta^I_J} B \quad C \xrightarrow{\delta^K_L} D}{B \otimes C \xrightarrow{\delta^{IK}_{JL}} A \otimes D}\ \backslash$$

In the full version of this paper (arXiv:1711:11513) we show that this labelling is correct for the general interpretation of proofs in §3.1.

3.2 Lexical semantics

For the general interpretation of types and proofs given above, a proof $f : A \longrightarrow B$ is interpreted as a linear map $\lceil f \rceil$ sending an element belonging to $\lceil A \rceil$, the semantic space interpreting A, to an element of $\lceil B \rceil$. The map is expressed at the general level of types, and completely abstracts from *lexical* semantics. For the computation of concrete interpretations, we have to bring in the meaning of the lexical items. For $A = A_1 \otimes \cdots \otimes A_n$, this means applying the map $\lceil f \rceil$ to $\mathbf{w}_1 \otimes \cdots \otimes \mathbf{w}_n$, the tensor product of the word meanings making up the phrase under consideration, to obtain a meaning $M \in \lceil B \rceil$, the semantic space interpreting the goal formula.

With the index notation introduced above, $\lceil f \rceil$ is expressed in the form of a generalised Kronecker delta, which is applied to the tensor product of the word meanings in index notation to produce the final meaning in $\lceil B \rceil$. In (4) we illustrate with the interpretation of some proofs derived from the same axiom leaves, $np \longrightarrow np$ and $s \longrightarrow s$. Assuming $\lceil np \rceil = \mathsf{N}$ and $\lceil s \rceil = \mathsf{S}$, these correspond to identity maps on N and S. We use the convention that the formula components of the endsequent are labelled in alphabetic order; the correct indexing for the Kronecker delta is obtained by working back to the axiom leaves.

$$\begin{array}{lllll}
a & & \mathsf{dream}^{np\backslash s} \longrightarrow np\backslash s & & \mathbf{dream}^{\mathsf{N}\otimes\mathsf{S}}_{i,j} \xrightarrow{\delta^{k,j}_{i,l}} T^{\mathsf{N}\otimes\mathsf{S}}_{k,l} \\
b & \mathsf{poets}^{np} \otimes \mathsf{dream}^{np\backslash s} \longrightarrow s & & \mathbf{poets}^{\mathsf{N}}_i \otimes \mathbf{dream}^{\mathsf{N}\otimes\mathsf{S}}_{j,k} \xrightarrow{\delta^{i,k}_{j,l}} V^{\mathsf{S}}_{l} & (4) \\
c & & \mathsf{poets}^{np} \longrightarrow s/(np\backslash s) & & \mathbf{poets}^{\mathsf{N}}_i \xrightarrow{\delta^{i,l}_{k,j}} R^{\mathsf{S}\otimes\mathsf{N}\otimes\mathsf{S}}_{j,k,l}
\end{array}$$

(4)(a) expresses the linear map from $\mathbf{dream} \in \mathsf{N} \otimes \mathsf{S}$ to a tensor $T \in \mathsf{N} \otimes \mathsf{S}$. Because we have $T = \delta^{k,j}_{i,l}\mathbf{dream}_{i,j} = \mathbf{dream}_{k,l}$, this is in fact the identity map. (4)(b) computes a vector $V \in \mathsf{S}$ with $V = \delta^{i,k}_{j,l}\mathbf{poets}_i \otimes \mathbf{dream}_{j,k} = \mathbf{poets}_j \otimes \mathbf{dream}_{j,l}$. In (4)(c) we arrive at an interpretation $R \in \mathsf{S} \otimes \mathsf{N} \otimes \mathsf{S}$ with $R = \delta^{i,l}_{k,j}\mathbf{poets}_i = \delta^{l}_{j}\mathbf{poets}_k$. Note that we wrote the tensor product symbol $\otimes$ explicitly.

In the case of our relative clause example (1), the derivational ambiguity of (3) gives rise to two ways of obtaining a vector $\mathbf{v} \in \mathsf{N}$. They differ in whether l, the index of the $\Diamond\Box np$ hypothesis in the relative pronoun type, contracts with index p for the subject argument of the verb (5) or with the direct object index o (6).

$$\begin{aligned} & \mathbf{mannen}_i \otimes \mathbf{die}_{jklm} \otimes \mathbf{vrouwen}_n \otimes \mathbf{haten}_{opq} \xrightarrow{\delta^{i,k,l,m,n}_{j,r,p,q,o}} \mathbf{v}^{subj}_r \in \mathsf{N} \\ \mathbf{v}^{subj}_j = \; & \mathbf{mannen}_i \otimes \mathbf{die}_{ijkl} \otimes \mathbf{vrouwen}_m \otimes \mathbf{haten}_{mkl} \qquad \text{(relabeled)} \end{aligned} \tag{5}$$

$$\begin{aligned} & \mathbf{mannen}_i \otimes \mathbf{die}_{jklm} \otimes \mathbf{vrouwen}_n \otimes \mathbf{haten}_{opq} \xrightarrow{\delta^{i,k,l,m,n}_{j,r,o,q,p}} \mathbf{v}^{obj}_r \in \mathsf{N} \\ \mathbf{v}^{obj}_j = \; & \mathbf{mannen}_i \otimes \mathbf{die}_{ijkl} \otimes \mathbf{vrouwen}_m \otimes \mathbf{haten}_{kml} \qquad \text{(relabeled)} \end{aligned} \tag{6}$$

The picture in Figure 3 expresses this graphically.

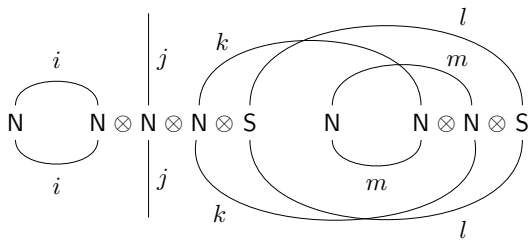


Figure 3: Matching diagrams for Dutch derivational ambiguity. Object relative (top), $\mathsf{mannen}_i\ \mathsf{die}_{ijkl}\ \mathsf{vrouwen}_m\ \mathsf{haten}_{kml}$ versus subject relative (bottom) $\mathsf{mannen}_i\ \mathsf{die}_{ijkl}\ \mathsf{vrouwen}_m\ \mathsf{haten}_{mkl}$.

Open class items vs function words For open class lexical items, concrete meanings are obtained distributionally. For function words, the relative pronoun in this case, it makes more sense to assign them an interpretation independent of distributions. To capture the intersective interpretation of restrictive relative clauses, Sadrzadeh et al [9] propose to interpret the relative pronoun with a map that extracts a vector in the noun space from the relative clause body, and then combines this by elementwise multiplication with the vector for the head noun. Their account depends on the identification $\lceil np \rceil = \lceil n \rceil = \mathsf{N}$: noun phrases and simple common nouns are interpreted in the same space; it expresses the desired meaning recipe for the relative pronoun with the aid of (some of) the Frobenius operations that are available in a compact closed category:

$$\Delta : A \to A \otimes A \quad \mu : A \otimes A \to A \quad \iota : A \to I \quad \zeta : I \to A \tag{7}$$

In the case of **FVect**, Δ takes a vector and places its values on the diagonal of a square matrix, whereas μ extracts the diagonal from a square matrix. The ι and ζ maps respectively sum the coefficients of a vector or introduce a vector with the value 1 for all of its coefficients.

$$\begin{array}{llccc}
\Delta_V : V \to V \otimes V & \text{given by} & \sum_i v_i \vec{e}_i & \mapsto & \sum_i v_i (\vec{e}_i \otimes \vec{e}_i) \\
\iota_V : V \to \mathbb{R} & \text{given by} & \sum_i v_i \vec{e}_i & \mapsto & \sum_i v_i \\
\mu_V : V \otimes V \to V & \text{given by} & \sum_{ij} v_{ij} (\vec{e}_i \otimes \vec{e}_j) & \mapsto & \sum_i v_{ii} \vec{e}_i \\
\zeta_V : \mathbb{R} \to V & \text{given by} & \lambda & \mapsto & \sum_i \lambda \vec{e}_i
\end{array}$$

The analysis of [9] uses a pregroup syntax and addresses relative clauses in English. It relies on distinct pronoun types for subject and object relativisation. In the subject relativisation case, the pronoun lives in the space $\mathsf{N} \otimes \mathsf{N} \otimes \mathsf{S} \otimes \mathsf{N}$, corresponding to $n^r\, n\, s^l\, np$, the pregroup translation of a Lambek type $(n\backslash n)/(np\backslash s)$; for object relativisation, the pronoun lives in $\mathsf{N} \otimes \mathsf{N} \otimes \mathsf{N} \otimes \mathsf{S}$, corresponding to $n^r\, n\, np^{ll}\, s^l$, the pregroup translation of $(n\backslash n)/(s/np)$.

For the case of Dutch, the homomorphism $\lceil \cdot \rceil$ of §3.1 sends the relative pronoun type $(n\backslash n)/(\Diamond\Box np\backslash s)$ to the space $\mathsf{N}\otimes\mathsf{N}\otimes\mathsf{N}\otimes\mathsf{S}$. This means we can import the pronoun interpretation for that space from [9], which now will produce both the subject and object relativisation interpretations through its interaction with the derivational semantics.

$$\textbf{die} \quad = \quad (1_{\mathsf{N}} \otimes \mu_{\mathsf{N}} \otimes 1_{\mathsf{N}} \otimes \zeta_{\mathsf{S}}) \circ (\eta_{\mathsf{N}} \otimes \eta_{\mathsf{N}}) \tag{8}$$

Intuitively, the recipe (8) says that the pronoun consists of a cube (in $\mathsf{N} \otimes \mathsf{N} \otimes \mathsf{N}$) which has 1 on its diagonal and 0 elsewhere, together with a vector in the sentence space S with all its entries 1. Substituting this lexical recipe in the tensor contraction equations of (5) and (6) yields the desired final semantic values (9) and (10) for subject and object relativisation respectively. We write $\odot$ for elementwise multiplication; the summation over the S dimension reduces the rank-3 $\mathsf{N} \otimes \mathsf{N} \otimes \mathsf{S}$ interpretation of the verb to a rank-2 matrix in $\mathsf{N} \otimes \mathsf{N}$, with rows for the verb's object, columns for the subject. This matrix is applied to the vector **vrouwen** either forward in (10), where 'vrouwen' plays the subject role, or backward in (9) before being elementwise multiplied with the vector for mannen.

$$(5) \quad = \quad \textbf{mannen} \odot \Big[\Big(\sum_S \textbf{haten} \Big)^{\mathsf{T}} \textbf{vrouwen} \Big] \tag{9}$$

$$(6) \quad = \quad \textbf{mannen} \odot \Big[\Big(\sum_S \textbf{haten} \Big) \textbf{vrouwen} \Big] \tag{10}$$

Returning to English, notice that the pregroup type assignment $n^r\, n\, np^{ll}\, s^l$ for object relativisation in [9] is restricted to cases where the 'gap' in the relative clause body occupies the final position. To cover these non-subject relativisation patterns in general, also with respect to positions internal to the relative clause body, we would use an $\mathbf{NL}_\diamond$ type $(n\backslash n)/(s/\Diamond\Box np)$ for the pronoun, together with the rightward extraction postulates $\alpha^r_\diamond, \sigma^r_\diamond$ of Figure 1. For English subject relativisation, the simple pronoun type $(n\backslash n)/(np\backslash s)$ will do, as this pattern doesn't require any structural reasoning.

4 Discussion

We briefly compare the distributional and the formal semantics accounts, highlighting their similarities. In the formal semantics account, the interpretation homomorphism sends syntactic types to their semantic counterparts. Syntactic types are built from atoms, for example s, np, n for sentences, noun phrases and common nouns; assuming semantic atoms e, t and function types built from them, one can set $\lceil s \rceil = t$, $\lceil np \rceil = e$, $\lceil n \rceil = e \to t$, and $\lceil A/B \rceil = \lceil B\backslash A \rceil = \lceil B \rceil \to \lceil A \rceil$. Each semantic type A is assigned an interpretation domain D_A, with $D_e = E$, for some non-empty set E (the discussion domain), $D_t = \{0, 1\}$ (truth values), and $D_{A\to B}$ funtions from D_A to D_B.

In this setup, a syntactic derivation $A_1, \ldots A_n \Rightarrow B$ is interpreted by means of a linear lambda term M of type $\lceil B \rceil$, with parameters x_i of type $\lceil A_i \rceil$ — linearity resulting from the fact that the syntactic source doesn't provide the copying/deletion operations associated with the structural rules of Contraction and Weakening.

As in the distributional model discussed here, the proof term M is an instruction for meaning assembly that abstracts from lexical semantics. In (11) below, one finds the proof terms for English subject (a) and object (b) relativisation. The parameter w stands for the head noun, f for the verb, y and z for its object and subject arguments; parameter x for the relative pronoun has type $(e \to t) \to (e \to t) \to e \to t$.

$$\begin{array}{ll} (a) & n, (n\backslash n)/(np\backslash s), (np\backslash s)/np, np \Rightarrow n \qquad (x_{who}\ \lambda z^e.(f^{e\to e\to t}\ y^e\ z^e)\ w^{e\to t}) \\ (b) & n, (n\backslash n)/(s/np), np, (np\backslash s)/np \Rightarrow n \qquad (x_{who}\ \lambda y^e.(f^{e\to e\to t}\ y^e\ z^e)\ w^{e\to t}) \end{array} \tag{11}$$

To obtain the interpretation of 'men who hate women' vs 'men who(m) women hate', one substitutes lexical meanings for the parameters of the proof terms. In the case of the open class items 'men', 'hate', 'women', these will be non-logical constants with an interpretation depending on the model. For the relative pronoun, we substitute an interpretation independent of the model, expressed in terms of the logical constant $\wedge$, leading to the final interpretations of (13), after normalisation.

$$x_{who} := \lambda x^{e\to t} \lambda y^{e\to t} \lambda z^e.((x\ z) \wedge ((y\ z)) \tag{12}$$

$$\begin{array}{ll} (a) & \lambda x.((\textsc{men}\ x) \wedge (\textsc{hate}\ \textsc{women}\ x)) \\ (b) & \lambda x.((\textsc{men}\ x) \wedge (\textsc{hate}\ x\ \textsc{women})) \end{array} \tag{13}$$

Notice that the lexical meaning recipe for the relative pronoun goes beyond linearity: to express the set intersection interpretation, the bound z variable is copied over the conjuncts of $\wedge$. By encapsulating this copying operation in the lexical semantics, one avoids compromising the derivational semantics. In this respect, the formal semantics account makes the same design choice regarding the division of labour between derivational and lexical semantics as the distributional account, where the extra expressivity of the Frobenius operations is called upon for specifying the lexical meaning recipe for the relative pronoun.

5 Acknowledgments

We thank Giuseppe Greco for comments on an earlier version. The second author would also like to thank Mehrnoosh Sadrzadeh for the many discussions on compositional distributional

semantics and Frobenius operations, and Rob Klabbers for his interesting remarks on index notation. The second author gratefully acknowledges support by a Queen Mary Principal's Research Studentship, the first author the support of the Netherlands Organisation for Scientific Research (NWO, Project 360-89-070, *A composition calculus for vector-based semantic modelling with a localization for Dutch*).

References

[1] Marco Baroni, Raffaela Bernardi, and Roberto Zamparelli. Frege in Space: a Program for Compositional Distributional Semantics. *Linguistic Issues in Language Technology*, 9:241–346, 2014.

[2] Bob Coecke, Edward Grefenstette, and Mehrnoosh Sadrzadeh. Lambek vs. Lambek: Functorial Vector Space Semantics and String Diagrams for Lambek Calculus. *Annals of Pure and Applied Logic*, 164(11):1079–1100, 2013.

[3] Joachim Lambek. On the Calculus of Syntactic Types. In Roman Jakobson, editor, *Structure of Language and its Mathematical Aspects*, volume XII of *Proceedings of Symposia in Applied Mathematics*, pages 166–178. American Mathematical Society, 1961.

[4] Joachim Lambek. Categorial and Categorical Grammars. In Richard T. Oehrle, Emmon Bach, and Deirdre Wheeler, editors, *Categorial Grammars and Natural Language Structures*, volume 32 of *Studies in Linguistics and Philosophy*, pages 297–317. Reidel, 1988.

[5] Jean Maillard, Stephen Clark, and Edward Grefenstette. A Type-Driven Tensor-Based Semantics for CCG. In *Proceedings of the Type Theory and Natural Language Semantics Workshop*, pages 46–54. EACL, 2014.

[6] Michael Moortgat. Multimodal Linguistic Inference. *Journal of Logic, Language and Information*, 5(3-4):349–385, 1996.

[7] Michael Moortgat. Constants of Grammatical Reasoning. In G Bouma, E Hinrichs, G.-J. Kruijff, and R.T. Oehrle, editors, *Constraints and Resources in Natural Language Syntax and Semantics*, pages 195–219. CSLI, 1999.

[8] Michael Moortgat and Richard Moot. Proof Nets for the Lambek-Grishin Calculus. In Chris Heunen, Mehrnoosh Sadrzadeh, and Edward Grefenstette, editors, *Quantum Physics and Linguistics: A Compositional, Diagrammatic Discourse*, pages 283–320. Oxford University Press, 2013.

[9] Mehrnoosh Sadrzadeh, Stephen Clark, and Bob Coecke. The Frobenius Anatomy of Word Meanings I: Subject and Object Relative Pronouns. *Journal of Logic and Computation*, pages 1293–1317, 2013.

[10] Gijs Wijnholds. Categorical Foundations for Extended Compositional Distributional Models of Meaning. *MSc. Thesis*, Universiteit van Amsterdam, 2014.

Lambdas, Vectors, and Word Meaning in Context

Reinhard Muskens[1] and Mehrnoosh Sadrzadeh[2]

[1] Tilburg University
r.a.muskens@gmail.com
[2] Queen Mary University of London
mehrnoosh.sadrzadeh@qmul.ac.uk

Abstract

We provide Lambda Logical Forms, which we think of as a reasonably neutral interface between syntax and semantics, with two interpretations. One interpretation is very close to a standard Montagovian semantics, be it that it allows the extensions of certain terms to be dependent on constants denoting vectors. A second interpretation constrains the values of these constants so that a form of word sense disambiguation in context results.

1 Introduction

Formal semantics and distributional semantics have complementary virtues. One approach explains how linguistic expressions can describe features of our surroundings, how a sentence can be true in one situation, but false in another, and what it means for expressions to be in a relation of entailment. The other approach provides a model of how the meanings of words can depend on other words and comes with a notion of similarity of meaning that often corresponds well with human judgements. One theory has a convincing story about the composition of phrasal meanings from pre-given word meanings, while the other actually provides those word meanings. One framework excels in the treatment of functional, especially logical, words, the other in the treatment of content words. And so forth.

How can the two approaches be combined? In previous work (Muskens and Sadrzadeh [8, 7]) we have shown how a Montague-like framework can be used to provide linguistic expressions—more precisely, abstract lambda terms that can stand proxy for linguistic expressions—with a vector-based semantics. Since the set-up made it possible to also provide those abstract lambda terms with a standard truth-functional semantics, a combination was obtained. But there was no communication between the two forms of semantics, which is clearly not satisfactory. In this paper we will remedy this by providing a set-up in which a truth-functional and a distributional component communicate through shared constants denoting vectors associated with word meaning in context.

2 Abstract Lambda Terms and Object Lambda Terms

Let us first explain the technical context that we assume here. It is clear that a semantic theory must be associated with a theory of syntax in order to be able to make predictions about the form-meaning relation in language. But there are many syntactic theories on the market and, in order to be able to avoid a choice between them, we will make use of some techniques from Abstract Categorial Grammars.[1] These will allow us to abstract away from the details of syntax and the details of various proposals for providing syntactic structures with a semantics. We

[1] De Groote [4]; see also Muskens [9, 10]. De Groote's Abstract Categorial Grammars (ACGs) and Muskens' Lambda Grammars were independently conceived in 2001, but are very similar. While ACGs can be used as a theory of surface syntax, this will not be the way we will use them in this paper.

constants	type
JOHN$_k$, SUE$_k$, MARY$_k$, ...	$(DS)S$
WOMAN$_k$, BALL$_k$, PARTY$_k$, FLU$_k$, ...	N
TALL$_k$, RED$_k$, STONE$_k$, ...	NN
SMOKE$_k$, TALK$_k$, RUN$_k$, ...	DS
LOVE$_k$, THROW$_k$, CATCH$_k$, ...	DDS
BELIEVE$_k$, CLAIM$_k$, HOPE$_k$, ...	SDS
WHO	$(DS)NN$
EVERY, A, NO, THE, MOST, ...	$N(DS)S$
AND, OR	$(\vec{\alpha}S)(\vec{\alpha}S)\vec{\alpha}S$

Table 1: Abstract constants (for each $k \in \mathbb{N}$) used for generating Lambda Logical Forms. AND and OR are assigned to each type of the form $(\vec{\alpha}S)(\vec{\alpha}S)\vec{\alpha}S$, where $\vec{\alpha}$ is any sequence of types.

will work with a level of abstract lambda terms that can be associated with syntactic structures in any of the usual ways but can also have various more concrete interpretations, as will be explained shortly.

The typed lambda terms that will form this interface of abstract terms will be called Lambda Logical Forms (LLFs). The types of these LLFs are defined to be the smallest set of strings such that (a) D, N, and S are (basic) types[2] and (b) whenever α and β are types, $(\alpha\beta)$ is a type.[3] We consider lambda terms over this set of types.[4] A *combinator* will be a closed lambda term not containing constants and a lambda term is called *linear* if every binder λX in it binds exactly one X. In Table 1 we have given a collection of constants, many of which must be subscripted with some $k \in \mathbb{N}$. We are interested in *occurrences* of content words, and each distinct occurrence of a same content word will be associated with a constant carrying a unique subscript. The set of LLFs is defined as the smallest set such that the following hold.

- Every constant in Table 1 is an LLF of the type(s) it is associated with in that table;
- every typed linear combinator is an LLF;
- if M is an LLF of type $\alpha\beta$ and N is an LLF of type α, then (MN) is an LLF of type β, provided no subscript in M also occurs in N;
- if M is an LLF and M is λ-convertible to a linear term M', then M' is also an LLF.

As examples of LLFs, here are first some linear combinators. (We use ξ as a variable of type D, $\mathcal{P}$ as a variable of type DS, $\mathcal{R}$ as a variable of type DDS, and $\mathcal{Q}$ as a variable of type $(DS)S$).

[2] D will be the type of names, N the type of nominal phrases, and S the type of sentences.

[3] This gives types with lots of parentheses in them, but outer parentheses will never be written and nor will parentheses be written if they can be recovered by the rule that association is to the right (so that DDS, the type of transitive verbs, is short for $(D(DS))$, for example).

[4] We will use the notation for lambda terms that is standard in formal work (see Barendrecht [1], for example), but not, alas, in linguistic applications. This means we will write (MN) (not $M(N)$) for the result of applying M to N and use $(\lambda X.M)$ (not $\lambda X(M)$) for lambda-abstraction. Outer parentheses can be removed and MNO is short for $(MN)O$ (i.e. association is to the left). But we will often refrain from removing (all) parentheses. The reason is that the structure of lambda terms in official notation is often quite close to that of the linguistic expressions they formalise, much closer in fact than the linguistic notation. This concerns hierarchical (dominance) aspects of the terms, not aspects to do with the order in which constants appear (linear precedence).

(1) a. $\lambda\xi\lambda\mathcal{P}.\mathcal{P}\xi$

b. $\lambda\mathcal{R}\lambda\mathcal{Q}_o\lambda\xi_s.\mathcal{Q}_o(\lambda\xi_o.\mathcal{R}\xi_o\xi_s)$

c. $\lambda\mathcal{R}\lambda\mathcal{Q}_o\lambda\mathcal{Q}_s.\mathcal{Q}_o(\lambda\xi_o.\mathcal{Q}_s(\mathcal{R}\xi_o))$

d. $\lambda\mathcal{R}\lambda\xi_1\lambda\xi_2.\mathcal{R}\xi_2\xi_1$

The reader will recognise (1a) as 'Montague Raising' (applied to, say, J of type D, it will give $\lambda\mathcal{P}.\mathcal{P}J$ of type $(DS)S$[5]), while (1b) and (1c) are forms of 'Argument Raising' (Hendriks [5]). Not all linear combinators are meaning preserving type raisers, however. (1d), for example, will change a relation to its converse. If the language of LLFs is used as an arbitrary interface with syntax, i.e. as an interface that virtually any syntactic theory can connect with, there is no guarantee that application of linear combinators will be meaning-preserving.[6]

In (2) some examples of LLFs of type S are given.

(2) a. Every man loves a woman
$((\textsc{every man}_0)(\lambda\xi_s.((\textsc{a woman}_1)(\lambda\xi_o.\textsc{love}_2\ \xi_o\xi_s))))$

b. Every man loves a woman
$((\textsc{a woman}_0)\lambda\xi.((\textsc{every man}_1)(\textsc{love}_2\ \xi)))$

c. Every tall woman smokes
$((\textsc{every}(\textsc{tall}_0\ \textsc{woman}_1))\textsc{smoke}_2)$

d. Sue loves and admires a stockbroker
$(\textsc{a stockbroker}_0)\lambda\xi.\textsc{sue}_1(\textsc{and admire}_2\ \textsc{love}_3\ \xi)$

e. Bill admires but Anna despises every cop
$(\textsc{every cop}_0)\textsc{and}(\lambda\xi.\textsc{anna}_1(\textsc{despise}_2\ \xi))(\lambda\xi.\textsc{bill}_3(\textsc{admire}_4\ \xi))$

f. The witch who Bill claims Anna saw disappeared
$\textsc{the}(\textsc{who}(\lambda\xi.\textsc{bill}_0(\textsc{claim}_5(\textsc{anna}_1(\textsc{see}_2\ \xi))))\textsc{witch}_3)\textsc{disappear}_4$

The reader will hopefully agree that any syntactic theory that comes with a way to associate syntactic structures with some form of lambda-based semantics, can also be coupled with LLFs.

But LLFs must be given a further interpretation in order to be useful for semantics. We explain how such a further interpretation can be given with the help of *type* and *term homomorphisms* (De Groote [4]). If $\mathcal{B}$ is some set of basic types, we write $TYP(\mathcal{B})$ for the smallest set containing $\mathcal{B}$ such that $(\alpha\beta) \in TYP(\mathcal{B})$ whenever $\alpha, \beta \in TYP(\mathcal{B})$. A function η from types to types is said to be a *type homomorphism* if $\eta(AB) = (\eta(A)\eta(B))$, whenever $\eta(AB)$ is defined. It is clear that a type homomorphism η with domain $TYP(\mathcal{B})$ is completely determined by the values of η for types $\alpha \in \mathcal{B}$. For example, let $\mathcal{B} = \{D, N, S\}$, the set of basic types of our LLFs, and let γ be the type homomorphism with domain $TYP(\{D, N, S\})$ such that $\gamma(D) = e$, $\gamma(N) = est$, and $\gamma(S) = st$ (as usual, e is for entities, t is for truth values, and s is for possible worlds). Then $\gamma(NN) = (est)est$, $\gamma(DS) = est$, $\gamma(DDS) = eest$, $\gamma(SDS) = (st)est$, $\gamma(N(DS)S) = (est)(est)st$, etc.

A function ϑ from lambda terms to lambda terms is a *term homomorphism based on* η if η is a type homomorphism and, whenever M is in the domain of ϑ:

[5] In fact, in Table 1 we have categorised constants such as $\textsc{john}_k$ directly in the type $(DS)S$.

[6] See Muskens [10] for a set-up in which application of linear combinators does not lead to form-meaning mismatches, because permutations in syntax and semantics always occur in tandem.

constant c	type c	c°	type c°
JOHN_k	$(DS)S$	$\lambda P.P\mathbf{j}$	$(est)st$
WOMAN_k	N	$woman$	est
RED_k	NN	$\lambda P\lambda x.red\ x \wedge Px$	$(est)est$
RUN_k	DS	run	est
THROW_k	DDS	$throw$	$eest$
BELIEVE_k	SDS	$\lambda p\lambda x\lambda w.\forall w'(Bxww' \rightarrow pw')$	$(st)est$
WHO	$(DS)NN$	$\lambda P'\lambda P\lambda x\lambda w.P'xw \wedge Pxw$	$(est)(est)est$
EVERY	$N(DS)S$	$\lambda P'\lambda P\lambda w.\forall x(P'xw \rightarrow Pxw)$	$(est)(est)st$
A	$N(DS)S$	$\lambda P'\lambda P\lambda w.\exists x(P'xw \wedge Pxw)$	$(est)(est)st$
AND	$(\vec{\alpha}S)(\vec{\alpha}S)\vec{\alpha}S$	$\lambda R'\lambda R\lambda\vec{X}\lambda w.R'\vec{X}w \wedge R\vec{X}w$	
OR	$(\vec{\alpha}S)(\vec{\alpha}S)\vec{\alpha}S$	$\lambda R'\lambda R\lambda\vec{X}\lambda w.R'\vec{X}w \vee R\vec{X}w$	

Table 2: A term homomorphism $(\cdot)^\circ$ sending (some) LLFs to object terms.

- $\vartheta(M)$ is a term of type $\eta(\tau)$, if M is a constant of type τ;
- $\vartheta(M)$ is the n-th variable of type $\eta(\tau)$, if M is the n-th variable of type τ;
- $\vartheta(M) = (\vartheta(A)\vartheta(B))$, if $M \equiv (AB)$;
- $\vartheta(M) = \lambda y.\vartheta(A)$, where $y = \vartheta(x)$, if $M \equiv (\lambda x.A)$.

Note that this implies that $\vartheta(M)$ is a term of type $\eta(\tau)$, if M is a term of type τ.

Clearly, a term homomorphism ϑ with domain the set of LLFs is completely determined by the values $\vartheta(c)$ for LLF constants c.

Here is an example of how this can be used. In Table 2 we have (partially) defined a term homomorphism $(\cdot)^\circ$ based on γ sending LLF constants to certain translations. (The types of variables and constants used in this table are given in a footnote.[7]) For the moment, and since this is merely an example, the translation is not sensitive to the values of subscripts on constants at all.

If this definition is extended to all LLF constants, we will automatically also get translations of all complex LLFs. Consider the LLF in (2a), for example. It's translation image under $(\cdot)^\circ$ in (3a) is easily seen to be equal to (3b), since $(\cdot)^\circ$ is defined to be a term homomorphisms. But, in view of the translations of constants in Table 2 (and similar ones that we leave to the reader), this is equal to (3c), which reduces to (3d) in the usual way.

(3) a. $((\text{EVERY MAN}_0)\lambda\xi_s.(\text{A WOMAN}_1)(\lambda\xi_o.\text{LOVE}_2\ \xi_o\xi_s))^\circ$

b. $(\text{EVERY}^\circ\ \text{MAN}_0^\circ)\lambda x.(\text{A}^\circ\ \text{WOMAN}_1^\circ)(\lambda y.\text{LOVE}_2^\circ\ yx)$

c. $((\lambda P'\lambda P\lambda w.\forall x(P'xw \rightarrow Pxw))man)$
$\lambda x.((\lambda P'\lambda P\lambda w.\exists x(P'xw \wedge Pxw))woman)(\lambda y.love\, yx)$

d. $\lambda w.\forall x\,(man\, xw \rightarrow \exists y\,(woman\, yw \wedge love\, yxw))$

[7] Using $A : \tau$ as shorthand for 'term A is of type τ', we have, for variables: $x, y, z : e$, $P : est$, $p : st$, $w : s$, $R : \vec{\alpha}S$, $\vec{X} : \vec{\alpha}$. For constants: $\mathbf{j} : e$, $woman$, red, $run : est$, $throw : eest$, $B : esst$.

constant c	type c	$c^\bullet$	type $c^\bullet$
JOHN$_k$	$(DS)S$	$\lambda Z.Z\mathbf{john}$	$(VV)V$
WOMAN$_k$	N	$\mathbf{woman}$	V
RED$_k$	NN	$\lambda v.(\mathbf{red} \times v)$	VV
RUN$_k$	DS	$\lambda v.(\mathbf{run} \times v)$	VV
THROW$_k$	DDS	$\lambda uv.(\mathbf{throw} \times_2 u) \times v$	VVV
BELIEVE$_k$	SDS	$\lambda uv.(\mathbf{believe} \times_2 u) \times v$	VVV
WHO	$(DS)NN$	$\lambda Zv.v \dotplus (Zv)$	$(VV)VV$
EVERY	$N(DS)S$	$\lambda vZ.Z(\mathbf{every} \times v)$	$V(VV)V$
A	$N(DS)S$	$\lambda vZ.Z(\mathbf{a} \times v)$	$V(VV)V$
AND	$(\vec{\alpha}S)(\vec{\alpha}S)(\vec{\alpha}S)$	$\lambda R'\lambda R\lambda\vec{X}.R\vec{X} \dotplus R'\vec{X}$	
OR	$(\vec{\alpha}S)(\vec{\alpha}S)(\vec{\alpha}S)$	$\lambda R'\lambda R\lambda\vec{X}.R\vec{X} + R'\vec{X}$	

Table 3: A term homomorphism $(\cdot)^\bullet$ sending LLFs of type S to terms denoting vectors.

Applied to a constant $\mathbf{w}$ (the 'actual world') the term in (3d) provides the desired truth conditions: $\forall x\,(man\,x\mathbf{w} \to \exists y\,(woman\,y\mathbf{w} \wedge love\,yx\mathbf{w}))$. The reader may find it amusing to translate other LLFs in a similar fashion and to experiment with alternative definitions of LLFs and of the type and term homomorphisms used. Our main point for the moment is that, given a collection of LLFs, the only thing needed for providing them with a semantics is to define a type homomorphism, plus a term homomorphism based on it.

3 Vectors in Type Logic

Since we want to combine truth-conditional semantics with vector semantics and use lambdas for composition, we must have a type theory that is able to talk about vectors over some field. For this field we choose the reals, as is usual. In order to have the latter available, we need a basic type $\mathbb{R}$ and constrain the models under consideration in such a way that the objects of type $\mathbb{R}$ are real numbers. Additionally, constants such as $0 : \mathbb{R}$, $1 : \mathbb{R}$, $+ : \mathbb{RRR}$, $\cdot : \mathbb{RRR}$, and $< : \mathbb{RR}t$[8] must have their usual interpretation. Fortunately, the problem of axiomatising the reals has already been solved for us by Alfred Tarski, who in [11] discusses two sets of (second-order) axioms for the real numbers. Adopting Tarski's axioms will ensure that the domain $D_\mathbb{R}$ of type $\mathbb{R}$ will equal the reals in full models.[9]

Vectors can now be introduced as objects of type $I\mathbb{R}$, where I is interpreted as some finite index set. Think of I as a set of words; if a phrase is associated with a vector $\mathbf{v} : I\mathbb{R}$, $\mathbf{v}$ assigns a real to each word, which gives information about the company the phrase keeps.[10] We abstract from the order present in vectors here. Since $I\mathbb{R}$ will be used often, we will abbreviate it as V. Note that $II\mathbb{R}$, abbreviated as M, can be associated with the type of *matrices* and $III\mathbb{R}$, abbreviated as C, with the type of *cubes*. In general, $I^n\mathbb{R}$ will be the type of tensors of rank n.

We need a toolkit of functions combining vectors, matrices, cubes, etc. Here are some definitions. The following typographical conventions are used for variables: r is of type $\mathbb{R}$; v and u are of type V; i, j, and ℓ are of type I; and m and c are of types M and C respectively.

[8] Constants such as $+$, $\cdot$, and $<$ will be written between their arguments.

[9] In generalised models that are not full the domain $D_\mathbb{R}$ will contain nonstandard reals, but will still satisfy the first-order theory of the reals.

[10] For exposition, we will work with a single index type I. Alternatively, several index types might be considered, so that phrases of distinct categories are allowed to live in their own space.

Indices are written as subscripts—v_i is syntactic sugar for vi.

$$
\begin{aligned}
* &:= \lambda rvi.r \cdot v_i : \mathbb{R}VV \\
\dot{+} &:= \lambda vui.v_i + u_i : VVV \\
\odot &:= \lambda vui.v_i \cdot u_i : VVV \\
\times &:= \lambda mvi.\sum_j m_{ij} \cdot v_j : MVV \\
\times_2 &:= \lambda cvij.\sum_\ell m_{ij\ell} \cdot v_\ell : CVM
\end{aligned}
$$

The reader will recognise $*$ as scalar multiplication, $\dot{+}$ as pointwise addition, $\odot$ as pointwise multiplication, $\times$ as matrix-vector and $\times_2$ as cube-vector multiplication. Other relevant operations are easily defined.

4 An Aside: Vectors All the Way Up

As a further example of how a set of LLFs can be provided with a semantics, we provide our fragment with a vector semantics in which phrases of all categories are associated with vectors, or vector-based functions. The type homomorphism we employ will be the function γ_1, defined by $\gamma_1(D) = \gamma_1(N) = \gamma_1(S) = V$, i.e. names, common nouns, and sentences all go to vectors. A term homomorphism $(\cdot)^\bullet$ based on γ_1 is given in Table 3. In this table the constants **john**, and **woman** denote vectors (type V), while **red**, **run**, **every**, and **a** denote matrices (type M), while **throw** and **believe** denote cubes (type C).[11] The variable Z is of type VV here.

We now find consequences of our translation such as the ones in (4).

(4) a. $((\text{A WOMAN}_0)\lambda\xi.((\text{EVERY MAN}_1)(\text{LOVE}_2\ \xi)))^\bullet =$
$(\mathbf{love} \times_2 (\mathbf{a} \times \mathbf{woman})) \times (\mathbf{every} \times \mathbf{man})$

b. $(((\text{EVERY}(\text{TALL}_0\ \text{WOMAN}_1))\text{SMOKE}_2))^\bullet = \mathbf{smoke} \times (\mathbf{every} \times (\mathbf{tall} \times \mathbf{woman}))$

c. $((\text{A STOCKBROKER}_0)\lambda\xi.\text{SUE}_1(\text{AND ADMIRE}_2\ \text{LOVE}_3\ \xi))^\bullet =$
$((\mathbf{love} \times_2 (\mathbf{a} \times \mathbf{stockbroker})) \times \mathbf{sue}) \dot{+} ((\mathbf{admire} \times_2 (\mathbf{a} \times \mathbf{stockbroker})) \times \mathbf{sue})$

d. $((\text{EVERY COP}_0)\text{AND}(\lambda\xi.\text{ANNA}_1(\text{DESPISE}_2\ \xi))(\lambda\xi.\text{BILL}_3(\text{ADMIRE}_4\ \xi)))^\bullet =$
$((\mathbf{admire} \times_2 (\mathbf{every} \times \mathbf{cop})) \times \mathbf{bill}) \dot{+} ((\mathbf{despise} \times_2 (\mathbf{every} \times \mathbf{cop})) \times \mathbf{anna})$

e. $(\text{THE}(\text{WHO}(\lambda\xi.\text{BILL}_0(\text{CLAIM}_5(\text{ANNA}_1(\text{SEE}_2\ \xi))))\text{WITCH}_3)\text{DISAPPEAR}_4)^\bullet =$
$\mathbf{disappear} \times (\mathbf{the} \times (\mathbf{witch} \dot{+} ((\mathbf{claim} \times_2 ((\mathbf{see} \times_2 \mathbf{witch}) \times \mathbf{anna})) \times \mathbf{bill})))$

It is of course the question whether it will be possible to harvest all the vectors, matrices and cubes that are necessary to make such translations more than a theoretical exercise. And, if so, it will still be the case that only empirical testing can answer the question how well a model such as this one will actually do on given tasks (say predicting perceived similarity of sentences). But, given that LLFs can form the output of many syntactic frameworks (and many parsers), there is at least no *theoretical* hurdle that must be overcome if we want to associate linguistic structures with a vector semantics.

[11] Compare Baroni and Zamparelli [2].

constant c	type c	$c^\dagger$	type $c^\dagger$
JOHN$_k$	$(DS)S$	$\lambda P.P(\mathbf{jv}^k)$	$(est)st$
WOMAN$_k$	N	$woman\,\mathbf{v}^k$	est
RED$_k$	NN	$\lambda P\lambda x.red\,\mathbf{v}^k x \wedge Px$	$(est)est$
RUN$_k$	DS	$run\,\mathbf{v}^k$	est
THROW$_k$	DDS	$throw\,\mathbf{v}^k$	$eest$
BELIEVE$_k$	SDS	$\lambda p\lambda x\lambda w.\forall w'(Bxww' \rightarrow pw')$	$(st)est$
WHO	$(DS)NN$	$\lambda P'\lambda P\lambda x\lambda w.P'xw \wedge Pxw$	$(est)(est)est$
EVERY	$N(DS)S$	$\lambda P'\lambda P\lambda w.\forall x(P'xw \rightarrow Pxw)$	$(est)(est)st$
A	$N(DS)S$	$\lambda P'\lambda P\lambda w.\exists x(P'xw \wedge Pxw)$	$(est)(est)st$
AND	$(\vec{\alpha}S)(\vec{\alpha}S)\vec{\alpha}S$	$\lambda R'\lambda R\lambda\vec{X}\lambda w.R'\vec{X}w \wedge R\vec{X}w$	
OR	$(\vec{\alpha}S)(\vec{\alpha}S)\vec{\alpha}S$	$\lambda R'\lambda R\lambda\vec{X}\lambda w.R'\vec{X}w \vee R\vec{X}w$	

Table 4: Term homomorphism $(\cdot)^\dagger$. As $(\cdot)^\circ$, but with additional vector constants.

5 Vectors in Truth-conditional Semantics

The fragment given in the previous section provides linguistic phrases with a vector semantics, but not with one that could easily be used to make predictions about truth conditions or entailment. It can be combined with an interpretation such as $(\cdot)^\circ$, and then, for each LLF Λ, Λ will come with a truth-conditional interpretation Λ° and a vector interpretation $\Lambda^\bullet$, but there is no interaction between the two.

We now want to present a model in which two interpretations, a distributional and a truth-conditional one, interact. The distributional interpretation will help to constrain word senses in context, while the truth-functional homomorphism will work on the basis of the senses thus constrained. Our inspiration for the distributional part has been Erk and Padó [3], but we replace their *structured vector space model* with a term homomorphism and their direct computation of word senses by a constraint-based approach.

The idea is as follows. Many words come, not with one, but with a whole collection of possible meanings. For example, the internet version of Merriam-Webster's dictionary gives no less than 18 different senses of the word *throw* (*throw a party*, *throw a ball*, *throw a tantrum*, ...). We assume that words typically come with a finite number of senses, each represented by a prototypical vector (see Kartsaklis et al. [6] for a closely related idea). While the senses of a word are often obviously related, two senses of the same word may well be associated with entirely different extensions.

The linguistic question is now how language users choose between these senses and the technological question is how a machine could be persuaded to do it. We focus on the linguistic question. Part of its answer must be that words do not only come with senses, but that senses also come with selectional preferences for other senses and that there is some mechanism for satisfying these preferences in the best possible way.

Let us first provide a truth-conditional set-up with vector senses, to be constrained shortly. In Table 4 we have defined a term homomorphism $(\cdot)^\dagger$, based on the type homomorphism γ that $(\cdot)^\circ$ was also based on. In fact, $(\cdot)^\dagger$ is very close to $(\cdot)^\circ$, but we now make use of the subscripts on some of the abstract constants, using them as superscripts on certain vector constants (the constants $\mathbf{v}^k$, of type V) in the translation. Constants such as *woman*, and *red* now have one extra argument place in order to provide for these constants. In particular, $\mathbf{j}$ is of type Ve, *woman* and *red* are of type $Vest$, and *throw* is of type $Veest$. Applied to constants of type V,

constant c	$c^\ddagger$	type $c^\ddagger$
JOHN$_k$	$\lambda\mathcal{Z}.\mathcal{Z}\lambda u.(\mathsf{john}\,\mathbf{v}^k \wedge u = \mathbf{v}^k)$	$(bt)t$
GRAPEFRUIT$_k$	$\lambda u.(\mathsf{grapefruit}\,\mathbf{v}^k \wedge u = \mathbf{v}^k)$	b
RED$_k$	$\lambda\rho\lambda u.(\mathsf{red}\,\mathbf{v}^k \wedge \rho u \wedge d(u, \mathbf{h} \times \mathbf{v}^k) < \varepsilon_k)$	bb
RUN$_k$	$\lambda\rho.\exists u(\mathsf{run}\,\mathbf{v}^k \wedge \rho u \wedge d(u, \mathbf{s} \times \mathbf{v}^k) < \varepsilon_k)$	bt
THROW$_k$	$\lambda\rho_1\rho_2.\exists uu'(\mathsf{throw}\,\mathbf{v}^k \wedge \rho_1 u \wedge \rho_2 u'$ $\wedge\, d(u, \mathbf{o} \times \mathbf{v}^k) < \varepsilon_k \wedge d(u', \mathbf{s} \times \mathbf{v}^k) < \varepsilon'_k)$	bbt
BELIEVE$_k$	$\lambda q\rho.\exists u(q \wedge \mathsf{believe}\,\mathbf{v}^k \wedge \rho u \wedge d(u, \mathbf{s} \times \mathbf{v}^k) < \varepsilon_k)$	tbt
WHO	$\lambda\mathcal{Z}\rho u.\mathcal{Z}\rho \wedge \rho u$	$(bt)bb$
EVERY, A, THE	$\lambda\rho\mathcal{Z}.\mathcal{Z}\rho$	$b(bt)t$
AND	$\lambda\mathcal{R}'\mathcal{R}\vec{X}.\mathcal{R}\vec{X} \wedge \mathcal{R}'\vec{X}$	
OR	$\lambda\mathcal{R}'\mathcal{R}\vec{X}.\mathcal{R}\vec{X} \vee \mathcal{R}'\vec{X}$	

Table 5: A term homomorphism $(\cdot)^\ddagger$ sending LLFs of type S to statements describing vectors.

they result in terms of the types they were given previously, and the rest of the set-up is as in the $(\cdot)^\circ$ homomorphism.

We get identities such as the one in (5), with vector constants helping to denote senses. An expression such as $run\mathbf{v}^2$ is intended to point at a particular sense of running (is it water running? an animal running? are we running out of time?).

(5) a. $(\text{JOHN}_1\ \text{RUN}_2)^\dagger = run\mathbf{v}^2(\mathbf{jv}^1)$ $\quad [=_\eta \lambda w.run\mathbf{v}^2(\mathbf{jv}^1)w]$

b. $((\text{A PARTY}_1)(\lambda\xi.\text{MARY}_2(\text{THROW}^3\ \xi))^\dagger =$
$\lambda w.\exists y\,(party\,\mathbf{v}^1 yw \wedge \rightarrow throw\,\mathbf{v}^3 y(\mathbf{mv}^2)w))$

c. $((\text{A UNICORN}_1)(\lambda\xi.(\text{EVERY DOG}_2)(\text{CHASE}_3\ \xi)))^\dagger =$
$\lambda w.\exists y\,(unicorn\,\mathbf{v}^1 yw \wedge \forall x\,(dog\,\mathbf{v}^2 xw \rightarrow chase\,\mathbf{v}^3 yxw))$

In order to make this work, the denotations of the $\mathbf{v}^k$ must be constrained and we do this with the help of a second homomorphism $(\cdot)^\ddagger$, defined in Table 5, and based on a type homomorphism γ' with $\gamma'(S) = t$ and $\gamma'(D) = \gamma'(N) = Vt$. Since Vt (sets of vectors) will be used often, we abbreviate it as b. $(\cdot)^\ddagger$ does not generate the usual kind of semantic objects, its sole purpose is to associate each LLF of type S with a conjunction of two kinds of statements:

- Statements saying that a certain vector belongs to the set of prototypes associated with a certain word. Examples: $\mathsf{run}\,\mathbf{v}^2$, $\mathsf{john}\,\mathbf{v}^1$, etc. Here the first constant is always of type Vt (i.e. b) and the second of type V. [Warning: $\mathsf{run}\,\mathbf{v}^2$ should well be distinguished from $run\,\mathbf{v}^2$.]

- Statements expressing that the cosine distance between two vectors is less than a given value. Example: $d(\mathbf{v}^1, \mathbf{s} \times \mathbf{v}^2) < \varepsilon_2$, which can be glossed as 'the distance between $\mathbf{v}^1$ and $\mathbf{s} \times \mathbf{v}^2$, the subject vector of $\mathbf{v}^2$, is sufficiently small. Here $\mathbf{s}$ is a matrix (type M).

The entries in Table 5 also mention other conjuncts, such as applicatons ρu and identities $u = \mathbf{v}^k$, but these only have intermediate importance in derivations.

Here is a simple example.

(6) $(\text{JOHN}_1\ \text{RUN}_2)^\ddagger = \mathsf{run}\,\mathbf{v}^2 \wedge \mathsf{john}\,\mathbf{v}^1 \wedge d(\mathbf{v}^1, \mathbf{s} \times \mathbf{v}^2) < \varepsilon_2$

We leave it to the reader to verify that the statement in (6) is correct. Note that in a last derivation step the predicate logical fact can be used that $\exists u(u = \mathbf{v} \wedge \varphi)$ is equivalent to the result of substituting $\mathbf{v}$ for u in φ. This is a general way of getting rid of existential quantifiers and identity statements that will recur often in derivations on the basis of $(\cdot)^\ddagger$.

The description generated as the right-hand side of (6) states that $\mathbf{v}^1$ is a vector associated with John, $\mathbf{v}^2$ is one of the prototype vectors associated with running, and that the distance between the John vector and a things-that-can-run vector associated with $\mathbf{v}^2$ is small. The idea that words come with vectors standing for their selectional preferences is taken from Erk and Padó [3], who "compute the selectional preference vector for word b and relation r as the weighted centroid of seen filler vectors $\vec{v}_a$". Here we associate the selectional preference vector not just with the word, but with each of the word senses.

Note that the statement $d(\mathbf{v}^1, \mathbf{s} \times \mathbf{v}^2) < \varepsilon_2$ in the right-hand side of (6) puts a *mutual* constraint on the vectors $\mathbf{v}^1$ and $\mathbf{v}^2$. Depending on the value of ε_2, certain combinations of values for $\mathbf{v}^1$ and $\mathbf{v}^2$ may well be excluded. This may be used to exclude certain models from consideration.

Let us have a look at some more images of LLFs under $(\cdot)^\ddagger$.

(7) a. $((\text{A}(\text{RED}_1\ \text{GRAPEFRUIT}_2))(\lambda\xi.\text{MARY}_3(\text{THROW}^4\ \xi)))^\ddagger =$
$\mathsf{throw}\,\mathbf{v}^4 \wedge \mathsf{red}\,\mathbf{v}^1 \wedge \mathsf{grapefruit}\,\mathbf{v}^2 \wedge d(\mathbf{v}^2, \mathbf{h} \times \mathbf{v}^1) < \varepsilon_1 \wedge \mathsf{mary}\,\mathbf{v}^3$
$\wedge\, d(\mathbf{v}^2, \mathbf{o} \times \mathbf{v}^4) < \varepsilon_4 \wedge d(\mathbf{v}^3, \mathbf{s} \times \mathbf{v}^4) < \varepsilon'_4$

b. $((\text{A UNICORN}_1)(\lambda\xi.(\text{EVERY DOG}_2)(\text{CHASE}_3\ \xi)))^\ddagger =$
$\mathsf{chase}\,\mathbf{v}^3 \wedge \mathsf{unicorn}\,\mathbf{v}^1 \wedge \mathsf{dog}\,\mathbf{v}^2 \wedge d(\mathbf{v}^1, \mathbf{o} \times \mathbf{v}^3) < \varepsilon_3 \wedge d(\mathbf{v}^2, \mathbf{s} \times \mathbf{v}^3) < \varepsilon'_3$

c. $(\text{THE}(\text{WHO}(\lambda\xi.\text{BILL}_0(\text{CLAIM}_5(\text{ANNA}_1(\text{SEE}_2\ \xi))))\text{WITCH}_3)\text{DISAPPEAR}_4)^\ddagger =$
$\mathsf{disappear}\,\mathbf{v}^4 \wedge \mathsf{see}\,\mathbf{v}^2 \wedge \mathsf{witch}\,\mathbf{v}^3 \wedge \mathsf{anna}\,\mathbf{v}^1 \wedge d(\mathbf{v}^3, \mathbf{o} \times \mathbf{v}^2) < \varepsilon_2 \wedge d(\mathbf{v}^1, \mathbf{s} \times \mathbf{v}^2) < \varepsilon'_2$
$\wedge\, \mathsf{claim}\,\mathbf{v}^5 \wedge \mathsf{bill}\,\mathbf{v}^0 \wedge d(\mathbf{v}^0, \mathbf{s} \times \mathbf{v}^5) < \varepsilon_5 \wedge d(\mathbf{v}^3, \mathbf{s} \times \mathbf{v}^4) < \varepsilon_4$

Note that in (7a) it is the vector connected with *grapefruit* that is constrained not to be too far from the object vector of *throw*. This is because the entry for RED_k in Table 5 'picks up' the value for u from its head, after which it can be picked up by further functors. In (7a) the vector for *witch* is not only constrained by the object vector of *see*, but also by the subject vector of *disappear*.

How can we define a notion of consequence on the basis of $(\cdot)^\dagger$ and $(\cdot)^\ddagger$? If Λ_1 and Λ_2 are LLFs, when does Λ_2 follow from Λ_1? We define a notion of entailment that is based on models that minimise the sum of the ε_k and ε'_k occurring in $(\Lambda_1)^\ddagger$ and $(\Lambda_2)^\ddagger$. Let us say that a model M is a δ-*model* if, (a) for no k, $\varepsilon_k > 1$ or $\varepsilon'_k > 1$ holds in M, (b) for only a finite number of k, $\varepsilon_k \neq 0$ or $\varepsilon'_k \neq 0$ holds in M, and (c) the sum of the nonzero values for ε_k and ε'_k in M is δ. Λ_2 *follows from* Λ_1 if there is a δ such that the following conditions hold.

- There is a δ-model M satisfying both $(\Lambda_1)^\ddagger$ and $(\Lambda_2)^\ddagger$;
- for every $\delta' \leq \delta$ and every δ'-model M, if M satisfies $(\Lambda_1)^\ddagger$, $(\Lambda_2)^\ddagger$, and $(\Lambda_1)^\dagger\mathbf{w}$, then M satisfies $(\Lambda_2)^\dagger\mathbf{w}$.

In the last clause, $\mathbf{w}$ is a fixed but arbitrary constant of type s that may be thought of as the actual world.

The idea here is that entailment holds if it there is transmission of truth in all models where the sum of the ε_k and ε'_k values are sufficiently low. As many vectors as possible need to be excluded as values for the $\mathbf{v}^k$. There may be ties, of course, in the sense that not all $\mathbf{v}^k$ are provided with a unique value, even in the best models (those with lowest δ), in which case there may be true ambiguity.

6 Conclusion

We have provided Lambda Logical Forms, with two interpretations. One interpretation is very close to a standard Montagovian semantics, be it that it allows the extensions of certain terms to be dependent on constants denoting vectors. The vectors can be thought of as prototypes associated with word senses. A second interpretation constrains the values of these constants so that a form of word sense disambiguation in context results. Since cosine distance must be minimised, many potential word senses are discarded. The disambiguation can be made sensitive to linguistic structure and is not restricted to local linguistic context.

7 Acknowledgments

We wish to thank the referees of Amsterdam Colloquium 2017 for their useful and encouraging comments.

References

[1] H. Barendrecht. *The Lambda Calculus: its Syntax and Semantics*. North-Holland, 1981.

[2] Marco Baroni and Roberto Zamparelli. Nouns are vectors, adjectives are matrices: Representing adjective-noun constructions in semantic space. In *Proceedings of the 2010 Conference on Empirical Methods in NLP*, EMNLP '10, pages 1183–1193. ACL, 2010.

[3] K. Erk and S. Padó. A structured vector space model for word meaning in context. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 897–906, 2008.

[4] Ph. de Groote. Towards Abstract Categorial Grammars. In *Association for Computational Linguistics, 39th Annual Meeting and 10th Conference of the European Chapter, Proceedings of the Conference*, pages 148–155, Toulouse, France, 2001. ACL.

[5] H.L.W.H. Hendriks. *Studied Flexibility: Categories and Types in Syntax and Semantics*. PhD thesis, University of Amsterdam, 1993.

[6] Dimitri Kartsaklis, Mehrnoosh Sadrzadeh, and Stephen Pulman. Separating disambiguation from composition in distributional semantics. In *Proceedings of 17th Conference on Natural Language Learning (CoNLL)*. Sofia Bulgaria, 2013.

[7] R. Muskens and M. Sadrzadeh. Context Update for Lambdas and Vectors. In M. Amblard, Ph. de Groote, S. Pogodalla, and C. Retoré, editors, *Logical Aspects of Computational Linguistics: 9th International Conference*, pages 247–254. Springer, 2016.

[8] R. Muskens and M. Sadrzadeh. Lambdas and Vectors. Abstract presented at the ESSLLI 2016 workshop on Distributional Semantics and Linguistic Theory (DSALT), 2016.

[9] R. A. Muskens. Categorial Grammar and Lexical-Functional Grammar. In Miriam Butt and Tracy Holloway King, editors, *Proceedings of the LFG01 Conference, University of Hong Kong*, pages 259–279, Stanford CA, 2001. CSLI Publications. http://cslipublications.stanford.edu/LFG/6/lfg01.html.

[10] R. A. Muskens. Language, Lambdas, and Logic. In Geert-Jan Kruijff and Richard Oehrle, editors, *Resource Sensitivity in Binding and Anaphora*, Studies in Linguistics and Philosophy, pages 23–54. Kluwer, 2003.

[11] Alfred Tarski. *Introduction to Logic and to the Methodology of Deductive Sciences*. Dover Publications, 1946.

Integrating lexical-conceptual and distributional semantics: a case report. *

Tillmann Pross, Antje Roßdeutscher, Sebastian Padó, Gabriella Lapesa, and Max Kisselew

University of Stuttgart

Abstract

By means of a case study on German verbs prefixed with the preposition *über* ('over') we compare alternation-based lexical-conceptual and usage-based distributional approaches to verb meaning. Our investigation supports the view that when distributional vectors are rendered human-interpretable by approximation of their representation with its nearest neighbour words in the semantic vector space, they reflect conceptual commonalities between verbs similar to those targeted in lexical-conceptual semantics. Moreover, our case study shows that distributional representations reveal conceptual features of verb meaning that are difficult if not impossible to detect and represent in theoretical frameworks of lexical semantics and thus that a general theory of word meaning requires a combination and complementation of lexical and distributional methods.

1 Introduction

A general theory of lexical representation is key to a compositional theory of the meaning of supralexical linguistic expressions. On these premises, the present paper investigates the relation between two approaches to word meaning: alternation-based lexical-conceptual semantics and usage-based distributional semantics.

In theoretical linguistics, a widely adopted hypothesis that drives research in lexical semantics is that "syntactic properties of phrases reflect, in large part, the meanings of the words that head them" [7]. One way to represent these syntactically relevant components of meaning is to decompose a verb's meaning into a fixed set of primitive predicates and constants from a limited set of semantic types. Typically, verbs of the same semantic class have common substructures in their decompositions, e.g. all verbs of change of state involve a substructure with the primitive 'become', and in which a constant names the state (e.g. 'broken') filling the second argument of 'become'. But syntactic properties of phrases have been argued to reflect even more fine-grained distinctions among verbs. For example, to explain the grammaticality of verbs in the conative construction, i.e. *She cut at the bread* vs. **She broke at the bread*, it has been proposed that the relevant distinction is of a conceptual nature. In the terminology of [10], the relevant distinction is realized by a "narrow-range" lexical rule: *cut* is a verb of motion, contact and causation whereas *break* is a verb of pure causation. Consequently, the concepts of motion, contact and causation must be represented in the particular meaning of a verb in a way that syntax can be sensitive to. That is, syntactic evidence not only provides a characterization of the general "templatic" aspects of verb meaning but also of the narrow-range constraints on the usage of a particular verb. As [6] shows impressively, when we extend the search for such syntactically represented conceptual distinctions to a wider range of verbs and constructions,

*The research reported in this paper has been supported by a DFG grant to the projects B4 (Pross, Roßdeutscher) and B9 (Padó, Lapesa, Kisselew) of the Collaborative Research Centre 732 "Incremental Specification in Context."

a systematic and fine-grained lexical-conceptual classification of verb meaning can be induced. We refer to this particular alternation-based approach of verb meaning in the following as the lexical-conceptual structure (LCS) approach to verb meaning.

A popular computational approach to lexical semantics, namely distributional semantic models (DSMs), starts from the hypothesis that "words that occur in similar contexts tend to have similar meanings" [12]. Accordingly, the distribution of a word's contexts are considered central to the construction of a suitable meaning representation of that word. A DSM representation of the meaning of a word is typically a point in a high-dimensional vector space, where the dimensions of the vector correspond to context items, e.g. co-occurring words, and the coordinates of the vector are defined by the strength of these context items, e.g. co-occurrence counts. Contextual similarity then becomes proximity of word meanings in the vector space. The DSM approach to word meaning is often illustrated by appeal to intuitions like the following (see e.g. [3]): *football* is similar in meaning to *soccer* since many of the words surrounding instances of *football* within a contextual window of a sentence are the same as the words surrounding instances of *soccer*. Theories of verb meaning like the LCS framework have been related to DSM approaches of word meaning with so-called "structured" DSM models [1], where DSM representations are not harvested from an unstructured window of tokens surrounding a given word, but from the distribution of words in specific syntactic-semantic frames. When the semantic feature spaces of structured DSM representations of contextual similarity are input to supervised classification or unsupervised clustering algorithms, verb classes similar to those identified in the LCS framework can be induced, see e.g. [11] for a discussion of the relationship between contextual similarity and theoretically defined verb classes. Another relevant distinction regarding DSM models concerns the way in which they are constructed. In what follows, we refer to classical DSMs built by accumulating co-occurrence information from structured or unstructured data as "count"-DSMs, and to DSMs extracted with neural network architectures as "predict"-DSMs. At the quantitative level, count DSMs are high-dimensional while predict DSMs are low-dimensional. From a qualitative point of view, the dimensions of count-DSMs correspond to actual words, while the dimensions produced by predict-DSMs can be thought of as soft clusters of context items [8] that do not correspond to actual words. However, whether or not the dimensions of a DSM model correspond to an actual word is insofar irrelevant as the adequacy of DSM representations is traditionally not determined by inspection of the DSM representation by itself but rather by evaluating the adequacy of a DSM representation against a gold standard (or a "Downstream Task") for a given clustering or classification problem. But by focusing solely on the successful reproduction of a gold standard, [5] concludes from a case study on structured DSM classification of Italian verbs, one may miss the right goal because one may well reproduce a given gold standard of classification while still there is "little understanding of the meaning components, i.e. the semantic features, relevant to analyze verb meaning". Importantly, the same difficulties with respect to the identification of the conceptual building blocks of word meaning arises for theoretical approaches to word meaning like the LCS framework, as the identification of those conceptual elements involved in narrow-range lexical rules and the definition of semantically cohesive subclasses of verbs are the methodological blind spot of the LCS approach to verb meaning. For example, [13] argues that the assumption that contact and motion are required for a verb to enter the conative construction are "purely stipulative" and that "there is no explanation why verbs that express motion and contact – and not even all of them – should enter into the alternation to the exclusion of verbs that do not".

We address the question for the conceptual building blocks of word meaning by using the unstructured predict-DSM approach to word meaning not only as a tool to reproduce an already

established (human-crafted) gold standard but as way to explore previously unknown conceptual aspects of word meaning and thus as a genuine technique of lexical semantics on par with alternation-based approaches like the LCS framework. We show that when predict-DSM representations are rendered human-interpretable by approximation of the representation with its nearest neighbour words in the semantic vector space, the resulting characterization reflects conceptual commonalities between verbs similar to the narrow-range lexical rules of Pinker or Levin's semantically cohesive subclasses and in fact reveals conceptual features of verb meaning that are difficult if not impossible to detect and represent in frameworks of lexical semantics like LCS. We develop our argument by comparing a classification of 80 German verbs prefixed with the preposition *über* ('over') into semantically cohesive verb classes à la Levin with the output of an unsupervised clustering of the same set of *über*-verbs (section 2). Second, we argue that rendering DSM representations transparent is not only highly diagnostic for word meaning but even more so for more complex cases of meaning composition (section 3). Adopting an additive model of the composition of DSM representations, we show that rendering transparent the difference vector that results from subtracting the DSM representation of a base verb from the DSM representation of an *über* prefixed-verb reveals insights into the conceptual underpinnings and effects of the process of prefixation like meaning shifts which, although linguistically reflected, standardly escape the attention of lexical semanticists. Section 4 concludes.

2 Simple meaning spaces

The basic use of *über* ('over') is as a preposition with two distinct meanings. Depending on the aspectual class of the matrix verb, an *über*-PP can refer to the direction of the motion of an accusative reference object as in (1) or to the location of a dative reference object as in (2).

(1) Der Mann sprang über den Zaun.
the man jump over the.ACC fence
"The man jumped over the fence"

(2) Das Bild hing über der Tür.
the painting hang over the.DAT door
'The painting hung above the door'

German has a productive mechanism of word formation by affixation of prepositional elements like *über* to a base verb. In the following, we distinguish four lexical-conceptual classes of German *über*-affixed verbs by considering the participation of these verbs in locative alternations, the licensing of PP complements and case assignment. First, when *über* is affixed to a verb as in (3), the derived verb describes a movement ACROSS some obstacle. As (1) shows, a PP complement construction with *über* is licensed with motion verbs like *springen*.

(3) Der Mann übersprang den Zaun.
the man over-PRFX.jump the.ACC fence
'The man jumped over the fence'

Second, when *über* is affixed to change of possession verbs like *geben* ('to give'), the prefixed verb describes the *transfer* of an object x from A to B as in (4). The argument marked with dative case identifies the location at which the transferred object x ends up. No *über* PP-complement construction is possible with the base verb (5).

(4) Er übergab ihr den Brief
he over-PRFX.give her.DAT the.ACC letter
'He handed her over the letter'

(5) *Er gab den Brief über sie.
he give the letter over her

A third class of *über*-affixed verbs describes the APPLICATION of an object to another object as in (6-a). This class of APPLICATION verbs is distinguished from the ACROSS class by participation

in a locative alternation as in (6-a)/(6-b).

(6) a. Peter überklebte den Kratzer mit einem Aufkleber.
Peter over-PRFX.glue the scratch with a sticker.
'Peter over-pasted the scratch with a sticker'
b. Peter klebte den Aufkleber über den Kratzer.
Peter paste the sticker over.PREP the scratch.
'Peter pasted the sticker over the scratch'

Fourth, *über*-affixation of a verb can also be used to describe that the event denoted by the verb exceeds a certain contextual standard on a SCALE provided by the base verb, see (7). No PP-complementation with *über* is possible for the SCALE-class and the direct object receives accusative case (8).

(7) Er überbewertete die Aktie.
he over-PRFX.value the.ACC share
'He overvalued the share'

(8) *Er bewertete über die Aktie
he value over the share

We assigned up to 20 *über*-prefixed verbs to each of the four lexical-conceptual classes identified in the previous section and extracted distributional vectors with 300 dimensions for the *über*-prefixed verbs and their morphologically and semantically related base verbs using the CBOW model proposed by [9] with a symmetric 5-word window. The vectors were extracted from SdeWac [4], a web corpus created from a subset of the DeWaC corpus. It contains about 45m sentences selected to be well-formed sentences. We use an unstructured DSM because these models are the simplest possible ones, make the fewest assumptions, and we were interested in assessing the topic-oriented perspective that they provide (rather than the relationally-oriented perspective of structured DSMs). We then computed pair-wise cosine similarity between the distributional vectors. We then tried to establish a hierarchy among the computed pairwise similarities with the hierarchical agglomerative clustering algorithm from the SciPy package using the unweighted pair group method with arithmetic mean as linkage algorithm. Manual inspection of the hierarchy output by the clustering showed that our lexical-conceptual classification is reproduced fairly well in that the verbs from the TRANSFER class (t) in (9), the SCALE class (s) in (10), the ACROSS class (a) in (11) and the APPLICATION class (ap) in (12) are by and large grouped together hierarchically. Certainly, each of the clusters contains some outliers, but closer inspections shows that these outliers are mainly due to errors in the preprocessing or ambiguities. This is a remarkable result, insofar as the underlying DSM is unstructured, whereas in computational linguistics verb classes are standardly reproduced with structured DSMs.

(9) TRANSFER übergehen (pass s.th.over) (a); übereignen (convey) (t); überführen (lead across) (a); übernehmen (take over) (t); überlassen (let s.o. s.th. for use) (t); überantworten (pass repsonsibility) (t); übersenden (send) (t); übermitteln (transfer)(t); überreichen (hand over) (t); übergeben (hand over) (t); überweisen (trans-scribe) (t);

(10) SCALE überstimmen (outvote) (s); überrepräsentieren (overrepresent) (s); überspielen (copy) (t); überhören (miss s.th.) (a); überreizen (overexite) (s); überfordern (overstrain) (s); überstrapazieren (overstrain) (s); übertreiben (overdo) (s) ; übersteigern (surmount) (s); überzeichnen (make burlesque) (s); überdrehen (overwind) (s); überspitzen (exaggerate) (s); überhöhen (inflate) (s); überladen (overload) (s); überfrachten (overcharge) (s); überschätzen (overestimate) (s); überbewerten (overrate)(s); übersehen (overlook) (a); überwiegen (outweigh) (s); überbuchen (overbook) (s);

(11) ACROSS übersetzen (translate) (t); überliefern (pass down) (t); überschreiben (transfer) (t); überlesen (skip) (a); überblättern (page over) (a); überfliegen (fly across) (a); überarbeiten (overwork) (s); überschreiten (overstep) (a); übertreten (cross) (a); überspringen (jump over) (a); überschauen (survey, overlook) (a); überkreuzen (cross) (a);

(12) APPLICATION überhängen (cover by hanging s.th.) (ap); überstreuen (cover with sprinkles) (ap); überstäuben (cover with dust) (ap); übergießen (douse) (ap); übersprühen (cover by spraying) (ap); überstreichen (cover with paint) (ap); übermalen (cover by painting) (ap); überkleben (paste over)(ap); überziehen (cover with a coat) (s); übertünchen (cover with whitewash) (ap); überdecken (cover) (ap); überlagern (overlay, interfere) (ap); überbauen (build s.th. across s.th.) (a); überklettern (climb over) (a); überwachsen (overgrow) (ap); übersäen (reseed) (ap); überragen (tower above)(a);

But the clustering allowed for an even more interesting insight, as it gave rise to the additional fifth cluster in (13), where verbs which we classified differently in our lexical-conceptual approach are clustered together.

(13) OVERPOWER überrollen (overrun) (a); überrennen (overrun)(a); überschwemmen (flood, drown) (ap); überfluten (deluge) (ap); überfallen (attack) (s); überwältigen (overwhelm) (s); überkommen (be assailed by sth.) (trans); übermüden (overfatique) (s); überfahren (knock down) (a); überfressen (overeat) (s); überschütten (spill s.th. on s.o.) (ap); überhäufen (heap on) (ap);

If, as is customary in computational linguistics, the quality of the clustering would be measured in terms of predicting the gold standard provided by our four hand-crafted lexical-conceptual classes, then we would have to conclude from (13) that the parameter settings of our clustering algorithm should be revised to achieve a higher precision. But closer inspection of the verbs in the fifth cluster suggests that there may be another option to interpret the clustering result: Maybe the additional cluster did not come about by accident but identifies an additional class of *über*-verbs which we were not able to detect with the admittedly simplistic lexical-conceptual diagnostic tools we employed. Because predict-DSM representations cannot be assessed to find out whether the fifth cluster came about by accident (and thus the algorithm is wrong) or is semantically cohesive (and thus the gold standard is wrong) we approximated the vector representations of the *über*-verbs in the fifth cluster with their "nearest neighbours" (where proximity in space of two vectors is identified by their dot product as in [8]) to determine the ten words nearest in the semantic vector space to the target word. Consider the base verb *rennen* ('to run') (14) and the derived verb *überrennen* ('to overrun') (15).

(14) *rennen* (to run) BASE
springen.V schnappen.V zurennen.V hüpfen.V wegrennen.V schreien.V brüllen.V
jump snap towards-run hop run-away scream yell
schleichen.V aufspringen.V schreiend.A
creep jump-up screaming

(15) *überrennen* (to overrun) DERIVED
Horde.N belagern.V Truppe.N Übermacht.N Streitmacht.N einmarschieren.v
hord besiege troop superiority force invade
stürmen.V erobern.V besiegen.V umzingeln.V
assault conquer defeat surround

What the representation for *(über)rennen* shows, and this generalizes to the verbs that were clustered together in (13), is that these verbs were not clustered together by accident but rather because they share a common conceptual core. The *über*-prefixed verbs describe unforeseeable events of overpowering instances of (natural) forces exertion. Interestingly, nothing in the lexical semantics of *rennen* or *über* (at least according to the standards of lexical-conceptual semantics) indicates the possibility of such a meaning shift through *über*-prefixation. Although apparently trivial, the observation that the nearest neighbour characterizations which can render opaque DSM representations interpretable by humans encode a certain kind of lexical-conceptual knowledge in the sense of Levin and Pinker has not been made in the literature before. One reason for this may be, as already mentioned, that DSM representations are standardly evaluated with respect to a gold standard. Gold standards are tied to specific purposes and hypotheses, whereas what we aim at doing is exploratory work, i.e. to try to give a linguistic interpretation to the information encoded in a DSM. Moreover, making DSMs transparent indicates an advantage of using an unstructured DSM, because the nearest neighbours of a given vector are topical in nature and do not require similarity with regard to the fillers of specific syntactic positions (e.g. direct objects). In this manner, they capture more abstract and general conceptual features of the semantic space, as indicated e.g. by the verbs *belagern* ('to besiege') and *umzingeln* ('to surround') in (15).

3 Complex meaning spaces

We suggested in the previous section that DSM representations encode aspects of word meaning that are difficult to target by means of grammaticality judgements at the syntax-semantics interface as in the LCS-framework. What kind of observations are we to expect for the composition of DSM representations? To approach this question, we adopted an additive model of the composition of DSM representations [2], and represented the meaning shift that results from the composition of a base verb with its prefix by the difference between the base verb vector and the prefix verb vector. Using the same method of nearest neighbour approximation as in the previous section, we rendered transparent the "shift" vector that results from subtracting the DSM representation of a base verb from the DSM representation of the corresponding *über*-prefixed verb. Thus, we did not try to learn one general DSM representation of the prefix *über* (because a general DSM representation will smooth out the meaning of *über*) but calculated for each pair of observed base and derived verb the specific "'surplus" that *über* makes to the construction. We then investigated the question whether a general semantic function of *über*-prefixation can be induced from the idiosyncractic meaning that our additive model of DSM representations assigns to *über* in a specific construction. Consider first (16).

(16) *kleben* (to glue) BASE

aufkleben.V	ausschneiden.V	Klebeband.N	festkleben.V	**bekleben.V**	
glue.on.PRTC.glue	out.PRTC.cut	tape	fix.glue	be.PRXF.glue	
verkleben.V	tropfen.V	ankleben.V	bemalen.V	abwischen.V	
fix	drop	on.glue	be.PRFX.paint	wipe-off	

(17) *überkleben* (to cover) DERIVED

Aufkleber.N	**bekleben.V**	Plakat.N	Schriftzug.N	Aufschrift.N	**kleben.V**
sticker	be.PRXF.glue	poster	letters	label	glue
aufkleben.V	bedrucken.V	Aufdruck.N	prangen.V		
on.PRTC.glue	be-print	logo	be-respleshdent		

(18) *über* (over) SHIFT
vorgenommen.A Bundesarchiv.N Bürgerbegehren.N Rüstungsexport.N
planned federal-archive petition-referendum export-of-arms
Freiheitsstrafe.N Umbenennung.N erfolgt.A Kürzung.N staatlich.A irreführend.A
prison-punishment re-naming done short-cut state misleading
propagandistisch.A
propaganda

When there are some shared nearest neighbours of the base vector and the derived vector (indicated by the bold face neighbours in (16)/(17)), the shift vector is basically noise and the meaning of the derived verb is compositional. That is, the combination of the verb *kleben* and the prefix *über* yields the APPLICATION meaning predicted by our lexical-conceptual classification, in which the meaning of the prefix and the derived verb is the same as the meaning of the preposition and the base verb in the locative alternation, see (6). In contrast, *schauen* ('to look')/*überschauen* ('to survey') as in (19)-(21) constitute a prototypical example where there are no salient shared neighbours of the base and the derived vector, but where the derived vector shares salient neighbours with the shift-vector.

(19) *schauen* (to look) BASE
gucken.V starren.V anstarren.V anblicken.V blicken.V anschauen.V angucken.V
peer stare at.PRTC.stare look-at-so. look look-at-s.o. peer-at-s.o.
grinsen.V lächeln.V reinschauen.V
grin smile look-into-s.th

(20) *überschauen* (to survey) DERIVED
überblicken.V **Komplexität.N** **Tragweite.N** Gestirn.N Mannigfaltigkeit.N
survey complexity bearing luminary complexity
Einbildungskraft.N Ansehung.N **Gesamtzusammenhang.N** Materie.N
imagination reputation totality interstellar-matter
unüberschaubar.A
unmanagable

(21) *über* (over) SHIFT
Komplexität.N Berücksichtigung.N Folgewirkung.N **Gesamtheit.N**
complexity taking-into-account consequence totality
Verflechtung.N Umwelteinwirkung.N Beeinträchtigung.N **Tragweite.N**
interconnection environment-consequence impairment bearing
Funktionsträger.N Differenzierung.N
administrator differentiation

We propose that when the overlap in nearest neighbours is greater between derived and shift vector ((20)/(21)) than between base and derived vector ((19)/(20)), this indicates that the meaning of the derived verb is figurative and that the meaning of the prefix *über* and the base verb *schauen* in combination is different from the meaning these words have in isolation. We call such a meaning of a complex expression that cannot be reduced to the meanings of its constituents "holistic". Tellingly, in contrast to *überkleben* (17), the base verb *schauen* is not among the nearest neighbours of the derived verb *überschauen* (20). The holistic semantic effect of prefixing *schauen* with *über* is linguistically reflected in the ungrammaticality of the locative alternation with *überschauen*. Whereas the meaning of the base verb *schauen* licenses the realization of the Ground argument with a PP-complement (22-a) but not as the direct object

of a prefix-construction (22-b), the holistic meaning of the prefix verb *überschauen* licenses the Ground argument only as a direct object (23-b) but not as a PP complement (23-a).

(22) a. Der Mann schaute über die Stadt.
the man look over.PREP the city
'The looked over the city.'
b. ?Der Mann überschaute die Stadt.
the man over-PRFX.see the city
'The man overlooked the city.'

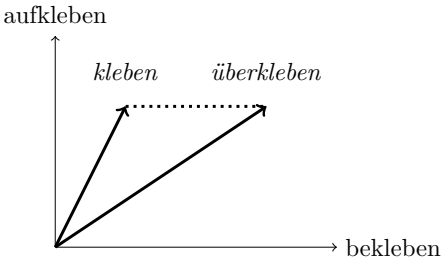
(23) a. *Der Mann schaute über die Komplexität des Problems.
the man look over the complexity the.GEN problem
b. Der Mann überschaute die Komplexität des Problems.
the man over-PRFX.look the complexity the.GEN problem
'The man surveyed the complexity of the problem.'

An intuitive explanation for the contrast between (22) and (23) may be given as follows. In (22), *schauen* is a perception verb that can be complemented with a PP specifying the perceptual space (i.e. that the subject has a view over the city). Consequently, because in (23) a spatial specification of the field of view with a PP is ungrammatical, this suggests that the relevant dimension of meaning in which *überschauen* is interpreted is no longer spatial, as would be expected for a verb that participates in the locative alternation. Instead, the composition of the verb and the prefix induces a holistic semantic effect by which the meaning of the prefix-verb is dislocated to a dimension of meaning not present in the prefix or the base verb in isolation. A quite similar holistic effect of meaning composition is involved in pure form in the fifth cluster of verbs of 'overpowering' (13), where the distributional characterization shows that the expected change of location reading is by and large replaced by the dislocated meaning of an unforeseeable event of (natural) force. In other words, whereas the meaning of the preposition and verb in the composition of *überkleben* is "rigid" (i.e. the meaning is not sensitive to context) and the salient dimensions of meaning of the preposition and the verb do not change through composition, the meaning of the preposition and verb in the composition of *überschauen* is 'non-rigid' and the salient dimensions of meaning of the preposition and the verb do change through composition. While such intuitions about the "dislocation" or "change" of a word's meaning dimensions are quite plausible when word meaning is perceived as a point in a high-dimensional vector space as DSM representations do, these intuitions are difficult to detect and represent in terms of lexical operations on the LCS of the base verb. Consequently, the way in which we phrased our intuitions hints towards the possibility that transparent DSM representations are better suited to make precise the semantic operation underlying the contrast between (22) and (23) on the one hand and *schauen* and *kleben* on the other. To foster an intuitive understanding of what it means that the meaning components denoted by the dimensions of a pair of vectors remain (mostly) unchanged in one case, but change in others, in the following we frame the contrast between (16)/(17) on the one hand and (22)/(23) on the other in a figurative understanding of meaning as a vector space. Thus, the following elaborations are neither intended as formally accurate explanations of DSM representations – in particular, we use nearest neighbours as approximations of dimensions – nor as lexical representations of word meaning in the traditional sense. Instead, we use the idea of meaning being represented in a vector space in a non-technical way to highlight what we believe is the specific "surplus" of DSM representations of meaning when compared against LCS-style analyses.

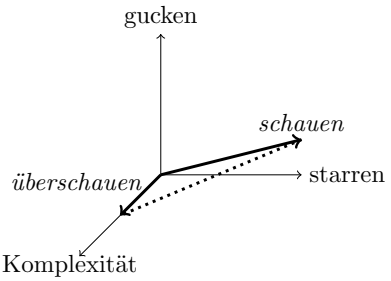
Consider first the simple rigid composition of *kleben* and *über*, where the base verb and the derived verb have salient nearest neighbours in common, i.e. the bold-faced nearest neighbours

in (16)/(17). For the sake of illustration, assume that we characterize the meaning of the base and derived verb with two of these shared salient nearest neighbours – (*bekleben* (paste sth. up) and *aufkleben* (to glue sth. on)) – and interpret the vectors associated with these neighbours as the dimensions of the meaning of the base and derived verb. Second, in the holistic case (19)/(20), the derived verb and the shift vector but not the base and derived verb share salient dimensions of meaning. Assume for the sake of illustration that we characterize the base verb *schauen* with its two most salient nearest neighbours *gucken* ('to peer') and *starren* ('to stare') and the derived verb with its most salient nearest neighbour *Komplexität* and that we use the vectors associated with these nearest neighbours as the meaning dimensions of the base and derived verb. The figures (24) and (25) visualize the meaning spaces characterized by these assumptions, where we represent the contribution of *über* according to our additive composition model as a dotted vector.

(24) rigid meaning composition



(25) holistic meaning composition



In (24) the meaning components denoted by the dimensions of the vectors remain (mostly) unchanged, but are deleted or overwritten in (25). That is, in (24) the composition of *über* and the base verb retains the original meaning dimensions and adds new dimensions already present in the meaning of the base verb, but in (25) the meaning dimensions of the base verb are replaced with new ones not present in the meaning of the base verb. Figuratively speaking, the derived verb *überkleben* lives in the same meaning space in which the base verb lives. In contrast, *überschauen* lives in a region of the meaning space different from that in which the constituents *überschauen* is composed of are located. In sum, whereas rigid composition is dimension-preserving and the meanings of *über* and *kleben* are the meanings these words have in isolation, holistic composition is non-dimension-preserving and the meaning composed of *über* and *schauen* cannot be decomposed to the meanings the preposition and the base verb have in isolation. Concluding, what we intend to make tangible with (24)/(25) is that the relation between lexical-conceptual semantics and DSM representations is more complex than it appears at first glance. In particular, the differences between the two are not just of a technical but also of a conceptual nature; the high dimensionality of the meaning space encoded in a DSM captures aspects of verb meaning that cannot be detected and represented with lexical frameworks like LCS (which focus on specific meaning dimensions like event or argument structure). But precisely because the "surplus" of DSM representations of word meaning falls outside the scope of traditional lexical semantics, this raises the question for how phenomena like the holistic meaning composition in (25) can be operationalized in a way that is compatible with established frameworks of lexical semantics like LCS. Given these complimentary strengths of LCS and DSM models of word meaning, we believe that a further investigation of the combination of lexical-conceptual and usage-based approaches may lead to an empirically grounded and theoretically sound theory of word meaning in its entirety.

4 Conclusion and Outlook

By means of a case study, we aimed to show that transparent DSM representations, when compared with the more traditional approach of lexical-conceptual semantics, provide a novel and exciting way to investigate the conceptual underpinnings of verb meaning in an empirically grounded and theoretically unbiased way. However, throughout the paper we were at pains to limit our attention to the discussion of observations we made rather than attempting to put forward a systematic theory of DSM representations and the principles of their composition. We remained reluctant with respect to broad claims about the nature and status of DSM representations because we simply put aside a question which, although of fundamental importance, we were not able to address given the goals and limitations of this paper. While it is standardly assumed in the literature (without further argument) that DSMs represent the meaning of words, in our case study we assumed that DSMs represent conceptual features (in the sense of Levin's cohesive semantics features of Pinker's narrow range lexical rules) only loosely associated with a specific word. In order to develop a systematic theory of what it is that DSM representations encode and consequently how DSM representations figure in the view of compositional meaning computation advanced in formal semantics, we believe that it is necessary to get a better understanding of what the objects of meaning are that DSM representations encode, for it makes a difference whether we are concerned with a theory of concepts and their linguistic expression or a theory of linguistic expressions and their conceptual underpinnings.

Bibliography

[**1**] M. Baroni and A. Lenci. Distributional memory: A general framework for corpus-based semantics. *Computational Linguistics*, 36(4):673–721, 2010.
[**2**] M. Baroni, R. Bernardi, and R. Zamparelli. Frege in space: A program of compositional distributional semantics. *LiLT*, 9:241 – 346, 2014.
[**3**] S. Clark. Vector space models of lexical meaning. In *The Handbook of Contemporary Semantic Theory*, pages 493 – 522. Wiley Blackwell, 2015.
[**4**] G. Faaß and K. Eckart. SdeWaC - a corpus of parsable sentences from the web. In *Proceedings of the International Conference of the GSCL*, 2013.
[**5**] A.Lenci. Carving verb classes from corpora. In *Word Classes: Nature, Typology and Representations*, p. 17 – 36. John Benjamins, 2014.
[**6**] B. Levin. *English verb classes and alternations: a preliminary investigation.* University of Chicago Press, 1993.
[**7**] B. Levin and S.Pinker. Introduction. In *Lexical & Conceptual Semantics*, p. 1–8. Blackwell, 1991.
[**8**] O. Levy and Y. Goldberg. Neural word embedding as implicit matrix factorization. In *Advances in Neural Information Processing Systems 27*, p. 2177–2185. Curran, 2014.
[**9**] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean. Distributed representations of words and phrases and their compositionality. In *Proc. of NIPS*, p. 3111–3119, 2013.
[**10**] S. Pinker. *Learnability and Cognition: The Acquisition of Argument Structure. New edition.* MIT Press, 2013.
[**11**] S. Schulte im Walde. Experiments on the Automatic Induction of German Semantic Verb Classes. *Computational Linguistics*, 32(2):159–194, 2006.
[**12**] P. D. Turney and P. Pantel. From frequency to meaning: Vector space models of semantics. *Journal of Artificial Intelligence Research*, 37:141–188, 2010.
[**13**] F. Van der Leek. The english conative construction: A compositional account. In *CLS 32*, p. 363–378, 1996.

The Formal Semantics of Free Perception in Pictorial Narratives *

Dorit Abusch and Mats Rooth

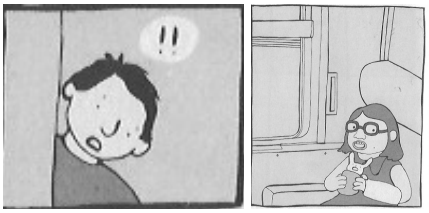
Cornell University, Ithaca, NY, USA

Abstract

This paper semantically analyzes "free perception" sequences in pictorial narratives such as comics, where one panel shows a character looking, and the next panel shows what they see. Pictorial contents are assumed to be viewpoint-centered propositions. A framework for the representation of pictorial narratives is used where indexing and embedding of certain panels is characterized by hidden operators. The resulting enriched pictorial narratives are interpreted in a dynamic framework. A possible worlds construction using action alternatives captures the epistemic effect of perceptual actions. Free perception sequences are implicitly anaphoric, as analyzed using cross-panel indexing. It is argued that some cases of free perception are truly intensional, and must involve embedding in the framework that is employed. Examples are drawn from comics and film.

1 Introduction

A common pattern in comics is a "free perception" sequence in which one panel shows a character looking, and the subsequent panel shows what is seen. The pair in (1a) is from S. Tan's *the Arrival*, showing a man looking down, and some enigmatic writing and graphics on the sidewalk.[1] It is understood that the second picture shows what the man sees. For another example, in Simone Lia's *Fluffy*, the character Michael has lost his rabbit Fluffy on a train. Searching, he looks into a cabin, and hallucinating, sees a girl eating a rabbit in a sandwich (see 1b). It is subsequently clarified that the girl was eating a kipferl, a kind of pastry.

(1) a.  b. 

The same phenomenon is found in film. (2) shows three frames from *the Third Man*, showing a man looking off camera to his left, with the final frame showing what he sees.[2]

*Thanks to Ede Zimmermann for comments. A preliminary version of this work was presented at Göthe University, Frankfurt in summer 2017. Thanks to the audience for their reactions. The images in the paper that are quoted from comics and film are used for educational and critical purposes, and are property of their respective owners.

[1] *The Arrival* is entirely wordless, lacking captions, thought bubbles, and speech bubbles. Such works are of special interest in the study of pictorial narratives.

[2] Such "eyeline match" transitions are part of the system of film continuity editing. Cumming et. al. (2017) is a semantic study of aspects of this system.

(2) 

There are closely similar examples in natural language narratives (Brinton 1980). Frequently they consist of an eventive clause that describes someone looking, followed by a stative clause describing what is seen. See (3a-c). Sometimes the information that a character looks is accommodated, as in (3d).[3]

(3) a. I looked back up the sidewalk, and that angry kid was walking toward me.
b. When I looked up a guy with a metal detector was walking toward me.
c. He looked at his mother. Her blue eyes were watching the cathedral quietly.
d. "Look!" Fred turned around. Jack was coming across the street towards him.

Current work on the semantics of pictures and pictorial narratives uses a possible-worlds model of information content (Greenberg 2011; Abusch 2012, 2016), based on the projective model of the semantic content of pictures (Hagen 1980). It is assumed here that a pictorial content is a viewpoint-centered proposition, modeled as a set of pairs of a world at a time and a geometric viewpoint (Rooth and Abusch 2017). A viewpoint is an oriented location in space, equivalent to the station point in the classical theory of perspective, or the location of an idealized camera. Functional notation is used for geometric projection, with $\pi(w, v, l, M) = p$ meaning that world-time w projects to picture p from viewpoint v. M and l are parameters for geometric projection.[4] Pictorial contents are obtained by inverting projection, $[\![p]\!]^{M,l} = \{\langle w, v\rangle | \pi(w, v, l, M) = p\}$.[5]

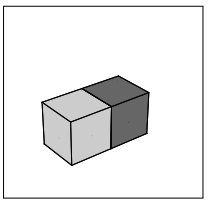
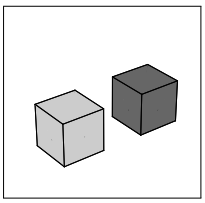
In order to model perceptual events and their epistemic properties, a construction of worlds as finite sequences of primitive events is assumed. Given a world w that satisfies the preconditions of an event a, wa is a world (at a time) where event a happened last. Perceptual events such as an agent looking come with event alternatives, and this is used in characterizing their epistemic properties. Thus we assume a construction of possible worlds as finite sequences of events, as in situation calculus (Reiter 2001), and a modeling of the epistemic consequences of events using Kripke relations on events, as in Baltag, Moss, and Solecki (1998).

Indexing across panels is significant in free perception sequences, because the agent about whom a free-perception picture gives visual-epistemic information is depicted in the previous panel. Characterizing the semantics of a free-perception panel involves reference to that agent, and this is a matter of indexing or anaphora across panels. Abusch (2012) introduced a syntactic approach to indices or discourse referents in pictorial narratives. Geometric points are interleaved with the narrative, and these points have the function of introducing and constraining model-theoretic values for discourse referents. Co-indexing is expressed with formal equalities. To illustrate, (5) is a short comic of two cubes moving apart, enriched with four discourse referents, and equalities between them. The notation is explained in a moment.

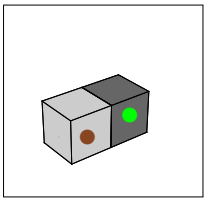
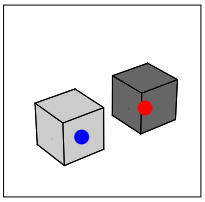
[3] (3a) is from a report by Larry Gross in CityBeat. (3b) is from the story "Ghosts" by Brian Hart. (3c) is from Lawrence's *Sons and Lovers*, as quoted by Brinton. (3d) is from Brinton (1980).

[4] l defines projection lines in terms of v, distinguishing for instance orthographic from perspectival projection. M is a marking rule that determines, for instance, that in (4), edges of geometric objects are marked in black.

[5] Abusch (to appear) is a survey of current work in this framework.

(4)  $(0.3, 0.3), (0.6, 0.4)$  $(0.3, 0.3), (0.7, 0.5), 1 = 3,\ 2 = 4$

An enriched pictorial narrative provides information about a world, a viewpoint, and a sequence of individuals, with the latter functioning as witnesses for discourse referents. (5) illustrates the form of a semantic satisfaction clause, where a certain tuple satifies a certain enriched pictorial narrative to the right of the turnstile. w is a world-state, constructed as above; v is a viewpoint, interpreted as the viewpoint for the last picture, and (x_1, x_2, x_3, x_4) is a tuple of witnesses for discourse referents. (In (5), the colors and colored dots are not part of the formula.)

(5) $w, v, (x_1, x_2, x_3, x_4) \models$  $(0.3, 0.3), (0.6, 0.4)$  $(0.3, 0.3), (0.7, 0.5), 1 = 3,\ 2 = 4$

Discourse referents are introduced with the interleaved geometric points. In (5), the point (0.7,0.5) is construed as a location in the preceding picture, and it introduces a discourse referent for the cube on the right in this picture (see the elements flagged in red). The point (0.3,0.3) introduces a discourse referent for the cube on the left in the last picture, flagged in blue. Similarly the points coming after the first picture introduce discourse referents for the cubes in that picture (flagged in green and brown). The semantics for discourse referents is random assignment, accompanied by a geometric constraint that locates objects in the model along a line determined by the current viewpoint and the geometric point specified in the discourse referent.[6] Formal equalities between natural numbers encode indexing across panels. A recency conventions is used: 1 is the most recently introduced discourse referent, 2 is the penultimately introduced discourse referent, and so forth. In (5), the equality 1=3 equates the dref for the cube on the right in the second picture with the dref for the cube on the right in the first picture. Similarly, 2=4 equates the drefs for the cubes on the left in the two pictures, which are flagged in blue and brown. The framework is comparable to a dynamic semantics for natural language where a discourse provides information about a world state and a list of individuals (Decker 2012).

The project for this paper is to use this toolkit to give a semantics for free perception in pictorial narratives. An important issue is the distinction between veridical free perception sequences such as (1a), where the free perception panel is construed as true of the base world timeline, and non-veridical ones such as (1b), where the base world timeline does not (or need not) satisfy the content of the free perception panel.

[6] See Abusch (to appear) for the details. Making it possible to state the semantics of discourse referents in this way is the motivation for storing the viewpoint for the last picture in the satisfying tuple.

2 LFs for free perception

We use the notation (p, q) for a free perception sequence, where p is the setup picture showing an agent looking, and q is the panel showing what the agent sees. In analyzing such sequences, there is an interplay between hypothesized logical forms for the narratives, interpretive principles for those logical forms, and modeling of the semantics of perceptual acts. We pursue a strategy of adding syntax to the narrative, in order to allow it to be interpreted incrementally and compositionally. Section 1 already mentioned that free perception sequences involve implicit anaphora to an agent in the first panel: a discourse referent for that agent should be added after the first panel, and then the semantics of the second panel should refer to that discourse referent. So a general hypothesis about the form of free perception sequences is (6), where p is the setup picture showing an agent looging, d introduces a discourse referent for that agent, and the complex $\phi(q, 1)$ interprets the second picture q in a way that explicitly or implicitly gives information about the visual-epistemic state of the agent. $\phi(q, 1)$ could involve syntactic embedding of q, or the addition of some conjuncts in a top-level sequence where q is a dynamic conjunct.

(6) $p\ d\ \phi(q, 1)$

To start, consider tuples that satisfy a non-enriched version of Fluffy sequence, as in (7a). Given the basic semantics, for any world w and viewpoint v that satisfy the narrative, $\pi(w, v, l, M) = q$, i.e. w looks like the second picture from viewpoint v. Of course, when we understand that Michael hallucinates, base worlds that satisfy the narrative do not (or need not) look like the rabbit sandwich picture from any viewpoint. The same point carries over to narratives with interleaved conjuncts. Whatever conjuncts are inserted in the position of the dots in (7b), any world w that satisfies the enriched narrative in the way shown in (7a) must have a prefix that satisfies the sandwich picture from some viewpoint. In other words, any narrative of the form seen in (7b) with the sandwich picture as a top-level conjunct entails (roughly) that a girl is eating or has eaten a rabbit sandwhich. This is the consequence of top-level pictures being interpreted extensionally, as providing information about what the base world (the world in the tuple to the left of the turnstile) looks like from some viewpoints at some times. Turning this result around, non-veridical free-perception panels are not top-level conjuncts.

(7) a. $w, v, \mathcal{O} \models \phi$  b. $w, v, \mathcal{O} \models \phi$  ...  ...

We deal with this conclusion by hypothesizing covert embedding of non-veridical free perception panels. The syntax in (8) is inspired by the syntax of clausal embedding in natural language. P is a covert verb (roughly, "see") that embeds the free perception panel as a complement, and has the index 1 as its covert subject. This index picks up the discourse referent for Michael that is introduced by d after the first panel. Given this syntax, it is the semantics of the phrase headed by P, rather than the sandwich picture, that places a constraint on the world variable to the left of the turnstile. This semantics is taken up in the next section. The syntactic proposal is fairly minimal, in that it gives access to the free-perception panel and the perceiving agent, and by embedding the free perception panel, it blocks an extensional interpretation.[7]

[7]The proposal is syntactic in the same way that the introduction of discourse referents and equalities between

(8) $w, v, \mathcal{O} \models \phi$  d 1 P 

What about free perception sequences that are understood veridically? In (1a) we understand that worlds that satisfy the narrative do look like the second panel from the visual perspective of the agent depicted in the first panel. And we understand that worlds consistent with *the Third Man* look like the third image in (2) from the perspective of the man depicted in the first two images. Should an embedding syntax as in (8) be used also for such cases? Or for them, should an extensional syntax be hypothesized? We develop both options.

The idea for an extensional analys of sequences such as (1) and (2) is that the free perception panel is a top-level conjunct, but with a particular geometric viewpoint enforced. In the satisfaction clause (9) v to the left of the turnstile memorizes the viewpoint for q (here p is the setup picture, d introduces a discourse referent for the agent in p, and q is the free-perception frame). The recursive semantics ensures that w looks like q from v. This viewpoint v is in principle unconstrained, but here is understood to be a geometric viewpoint determined by the agent 1, corresponding to the location of the eyes (or other visual system) of that agent. Accordingly we add a geometric predicate $V(x)$, which contributes the geometric constraint that the ambient viewpoint is the oriented location of x's visual system. When $V(1)$ is added to the right of the free perception panel as in (10), it enforces that the viewpoint for the free perception panel is the geometric visual viewpoint for agent 1. In this, both the panel q and the predication $V(1)$ are extensional.

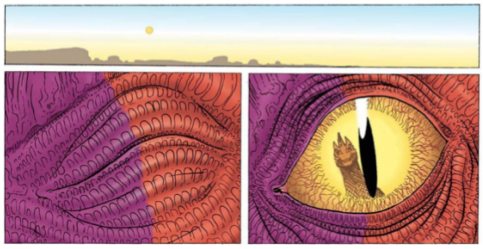
(9) $w, v, \mathcal{O} \models p\ d\ q$

(10) $w, v, \mathcal{O} \models p\ d\ q\ V(1)$

There are a couple of different panel types that are pragmatically similar to veridical free perception. (11a) is from Cece Bell's autobiographical *El Deafo*, and shows the heroine Cece and another character, Ginny. A dotted sightline indicates that Ginny is looking at Cece's hearing aid. Sightlines are a convention that indicate the visual focalization of a depicted agent. The information that is conveyed is quite similar to what would be conveyed by a free perception sequence, with one panel showing Ginny looking, and the next panel showing the hearing aid and the top of Cece's head. The information conveyed by (11a) appears to be entirely extensional—the characters are in a certain geometrical configuration, and Ginny is visually focalizing on a certain point. The panel carries the information that Ginny is looking, and gives information about what she is focalizing visually. But is arguably neutral about what information she picks up.

them is syntactic. In particular, it is the enriched narrative rather than the surface narrative that is interpreted compositionally. This way of proceeding is similar to what is seen in discourse representation theory (Kamp and Reyle 1993).

(11) a. 

b. 

(11b) is from Delgado's *the Age of Reptiles*. A predatory dinosaur opens its eye, and in the last panel, another dinosaur is seen reflected in the eye. It is inferred that the predatory dinosaur sees the other one, with an ominous implication that it has spotted its prey. But the literal information in the panel is extensional.

Consider for a moment what would be involved in an intensional syntax and interpretation for (11b). The panel would have to be broken down into two sub-panels, one showing the dinosaur looking, and another a small, syntactically embedded subpanel showing what is seen. This amounts to a "vision bubble" embedded in an image of the agent's eye. There are genuine vision bubbles, as seen in (12a) from Bilal's *Cold Equator*. But such an analysis is otiose in the case of (11)b, because of the possibility of a straightforward extensional interpretation.

(12b) is from Tezuka's *Ode to Kirihito*. It shows a hulking figure at a door, with his head tilted down towards the hero Kirihito on the floor. It can be inferred that the hulking figure sees a view approximately like the part of the panel surrounding Kirihito. But the panel as a whole could not show what the hulking character sees, because he himself is depicted. Here again an extensional analysis is attractive.

(12) a. 
b. 

These three panel types (with sight lines, eye reflections, and over-the-shoulder viewpoint) are pragmatically similar to veridical free perception. There is little temptation in these cases to formulate an intensional analysis based on a syntax with embedding, since the inferences that readers tend to make about what characters see are supported by the extensional content of the panel. This tends to favor an extensional analysis of veridical free perception, because here too (assuming a switch in geometric viewpoint as enforced in (10)), the inferences that we make about what the agent sees are supported by the extensional content of the sequence.

3 Models for misperception and veridical perception

This section defines a model of perception in the event framework sketched in Section 1. The main idea is to model veridical perception and mis-perception using alternatives to perceptual events. The relation of alternativeness is like an accessibility relation in a Kripke model for knowledge and belief modalities, except that it operates at the level of events, rather than

worlds. This way of proceeding is based on Baltag, Moss and Solecki (1999).

We use the term $l(x, p)$ to represent the event of agent x looking veridically at a scene that projects to picture p from x's geometric perspective. This is an atomic event, which in the way reviewed in Section 1 figures in the construction of possible worlds. Such events have a role as event *types*, in that event $l(x, p)$ can occur in different world-time lines, or be repeated in a single world timeline. The properties of $l(x, p)$ are captured by its pre-conditions, and by its visual-epistemic alternatives for agent x.

Preconditions in situation calculus are used to capture the physics and metaphysics of the modal space. The elevator can go down only if it is above the ground floor. Block b can be placed on block a only if block a has a clear top surface. In the possible worlds model, world w can be incremented with event e to form world we if and only if the preconditions of e are true in w.[8] We think of $l(x, p)$ as a highly specific event of looking, which can happen only in worlds w where agent x is facing a scene that looks like p from the agent's geometric perspective. The position and orientation of agent x in w depends on the world history w—how x has moved in this history. The highly specific looking act $l(x, p)$ can happen in w only if that history is such that at the world/time w, x is facing a scene that looks like picture p. If this precondition is met, there is an incremented world $w \cdot l(x, p)$, where x has just performed an act of veridical looking.[9]

Epistemic properties of events are captured with a relation of event-alternatives. For a perceptual event e, taking the alternative-set for e to be the unit set e provides a modeling of veridical looking. Consider world a $w \cdot l(x, p)$, where $l(x, p)$ has just happened. Arguably any world of the form $u \cdot l(x, p)$ is consistent with the visual-epistemic information in the event $l(x, p)$ that just happened in $w \cdot l(x, p)$. In particular, because of the precondition, in u agent x is facing a p-like scene. If looking does not change the geometric facts, this is true also in $u \cdot l(x, p)$. Veridicality amounts to $w \cdot l(x, p)$ itself being a world of the form $u \cdot l(x, p)$, meaning that x is also facing a p-like scene in the base world. The agent is facing a p-like scene in both the base world, and any visual-epistemic alternatives for the agent. On top of this, the event $l(x, p)$ has just happened in the base world, and in any visual-epistemic world alternative. This is a kind of introspection condition on the source of the visual-epistemic information.

Using Q_x for the perceptual-alternative relation for agent x, these ideas are recorded in (13).

(13) Visual-epistemic event alternatives for $l(x, p)$
$Q_x(l(x, p)) = \{l(x, p)\}$

Visual-epistemic world alternatives determined by $l(x, p)$
$\bar{Q}_x(l(x, p)) = \{u \cdot l(x, p) | u \text{ satisfies the preconditions of } l(x, p)\}$

This account distinguishes the visual-epistemic content of the looking event from the epistemic state of the agent after looking. A world $v \cdot l(x, p)$ can be consistent with the perceptual information in the looking event that has just happened in $w \cdot l(x, p)$, but inconsistent with x's overall information in $w \cdot l(x, p)$. Let R_x be the epistemic alternative relation for agent x. (14) gives a principle in deduction format for updating R_x when a world w is extended with a perceptual action e of x to form $w \cdot e$. It amounts to what was seen before, but with the alternative $v \cdot e'$ required to be formed from a world v that is an epistemic alternative for x in w.

[8] See Reiter (2001) for a development of these concepts.

[9] Normally w can be extended in other ways, for instance with an axtion $s(x)$ of the agent stepping forward. So this is a branching-time model.

(14) $\dfrac{R_x(w, v) \quad Q_x(w \cdot e, v \cdot e')}{R_x(w \cdot e, v \cdot e')}$

Discussions of free perception in language emphasize that it describes perceptual content, not epistemic state in the general sense (Kuroda 1976; Brinton 1980). Passage (3d) is understood to entail that Fred saw Jack coming across the street, not merely that he believed or knew he was. The same is true of pictorial free perception as analyzed using (13). $\bar{Q}_x(l(x, p))$ is a propositional content for the perceptual event $l(x, p)$, which is stated without reference to the epistemic state of the agent.

Veridical looking is characterized by visual-epistemic alternatives being similar to the base world in the way formalized in (13). In mis-perception, alternatives are not as similar to the base world. When Michael looks into the cabin, he sees a view q_r of a girl eating a rabbit sandwich. He believes he is engaged in veridical perception rather than mis-perception. This means his visual-epistemic world alternatives are of the form $v \cdot l(x, q_r)$, just as before. The difference is that the *base* world is not of this form. We introduce an additional basic looking action $m(x, p)$, thought of as an event of x looking at a scene which for x is p-like, but which is not (or is not necessarily) p-like in the base world.

(15) Visual-epistemic event alternatives for $m(x, p)$
$Q_x(m(x, p)) = \{l(x, p)\}$

Visual-epistemic world alternatives determined by $m(x, p)$
$\{v \cdot l(x, p) | v$ satisfies the preconditions of $l(x, p)\}$

For a simple idealized model, it is stipulated that events of the form $l(x, q)$ and $m(x, q)$ are the only looking events. A good setup panel and discourse referent for free perception is one which entails that the agent has just looked, i.e. that the last event that happened is either $l(x, q)$ or $m(x, q)$. or $w \cdot m(x, p)$. These are setup pictures where it "looks like" the agent picked out by the discourse referent is looking. We make the further assumption that actions of the form $l(x, p)$ are for the agent x alternatives only to looking actions. That is, if $l(x, p)$ is an element of $Q_x(e, l(x, p))$ then e is of the form $l(x, p')$ or $m(x, p')$.

Events $l(x, q)$ are used in scenarios of veridical looking, and events $m(x, q)$ are used in scenarios of mis-perception. Should it be assumed that events of the second kind are always erroneous, in the extensional sense that the base world does *not* look like q from x's geometric perspective? Consider a world w that looks like q from agent x's geometric perspective. World w satisfies the precondition of $l(x, q)$, and $w \cdot l(x, q)$ is a world where x has just looked veridically. If $w \cdot m(x, q)$ is also defined, then it is a formally different world which has the same visual-epistemic alternatives for x. So w branches into two worlds $w \cdot l(x, q)$ and $w \cdot m(x, q)$, that do not differ in properities that we want to model. This oddity is eliminated with a precondition for $m(x, q)$ that the world does not look like q from x's perspective (though the agent sees it as looking like q). We adopt this precondition for $m(x, q)$.[10]

[10] However, the other choice is also reasonable. If we think of $m(x, q)$ as x hallucinating a q-scene due to some specific effects in the low-level visual system or the cognitive system, it could be that x sees q due to those effects, but is accidentally right, in that x is facing a q scene in the base world. In this case, $l(x, p)$ happening should be distinguished from $m(x, p)$ happening, because only the first leads to knowledge. This comes up in Gettier scenarios.

4 Semantics for the LFs

This section interprets the LFs for free perception that where suggested in Section 2 in the event models from Section 3. (16) is the embedding LF, where q is embedded under P. The geometric point d sets up a discourse referent that can be referenced as $\mathcal{O}[1]$. ϕ is the part of the narrative preceding the free perception sequence.

(16) $w', v', \mathcal{O} \models \phi\ p\ d\ [1\ [P\ q]]$

Let w' be decomposed as $w \cdot e$, so that e is the event that just happened in w'. Where x is the agent $\mathcal{O}[1]$, $\overline{Q}_x(e)$ is the set of worlds that are perceptual alternatives to the event e that happens in the base world. Roughly, the semantics for the embedding construction should do a subset check between the visual-epistemic alternatives $\overline{Q}_x(e)$, and the content $[\![q]\!]^{M,l}$ of the embedded picture. Since content of the picture is viewpoint-centered, $\overline{Q}_x(e)$ needs to be adjusted to the viewpoint-centered proposition $\{\langle u', v'\rangle | u' \epsilon \overline{Q}_x(e) \wedge v' = \overline{V}(u', x)\}$, pairing the alternative world u' with the geometric viewpoint of x in u'. $\overline{V}$ is a function that maps a world and an agent to the geometric viewpoint of the agent in the world.[11] All of this leads to the semantics (17) for the embedding construction.

(17)
$$\frac{\begin{array}{c} w \cdot e, v, \mathcal{O} \models \phi \\ \mathcal{O}[n] = x \\ \{\langle u', v'\rangle | u' \epsilon \overline{Q}_x(e) \wedge v' = \overline{V}(u', x)\} \subseteq [\![q]\!]^{M,l} \end{array}}{w, v, \mathcal{O} \models \phi\ [n\ [P\ q]]}$$

A tricky question is what to do about the viewpoint in the conclusion. Normally a panel resets the viewpoint to the viewpoint from which the base world projects to the panel. In this case, since q is embedded, it is not projected in the base world, and there may be no viewpoint from which the base world projects to q. We have left the viewpoint constant.

On top of the truth conditions encoded in (17), it seems natural to say that $[n\ [P\ q]]$ presupposes that in w, $\mathcal{O}[n]$ is an agent with a visual system, and that e (the last event in $w \cdot e$) is a looking action by that agent. In the simple model construction where there are just two kinds of looking, $[n\ [P\ q]]$ presupposes that the base world finishes with either $l(x, q')$ or $m(x, q')$, for some q'.

(18) is the extensional option for the logical form of free perception. Herethere is nothing more to say about the semantics of q, since it is interpreted extensionally as placing a constraint on w and v. We just have to recall that $V(1)$ constrains v to be the geometric visual viewpoint of $\mathcal{O}[1]$, $v = \overline{V}(w, \mathcal{O}[1])$. This enforces that w looks like q from the geometric visual viewpoint of agent $\mathcal{O}[1]$.

(18) $w, v, \mathcal{O} \models p\ d\ q\ V(1)$

Section 2 finished with the question whether apparently veridical free perception sequences should be analyzed with the embedding LF (16), or with an LF where the free perception panel is in an extensional position as in (18). These options come out as symmetric in one dimension. The embedding LF expressed that things look like q for the agent, as expressed by the agent's visual-epistemic alternatives being of the form $u \cdot l(x, q)$. It presupposes that the agent is looking in the base world, but the base world could be either of the form $w'' \cdot l(x, q)$, with the agent facing a q scene in the base world, or of the form $w'' \cdot m(x, q')$, with q' not equal to q and the

[11] As Ede Zimmermann pointed out to us, it would be nice at this juncture if the alternatives were agent-centered worlds, rather than worlds. Then it would not be necessary to identify the agent across worlds.

agent facing some other kind of scene in the base world. The extensional LF (18) entails that w looks like q from viewpoint v. If w finishes with a looking event by x, it could finish either with $l(x, q)$, or with $m(x, q')$, with $q' \neq q$. Thus the embedding LF indicates what the agent's visual alternatives look like, and is neutral about what the base world is like. The extensional LF indicates what the base world looks like, and is neutral about what the agent's visual-epistemic alternatives look like. Resolving this issue requires further investication of what the entailments of examples such as (1a) should be.

5 Discussion

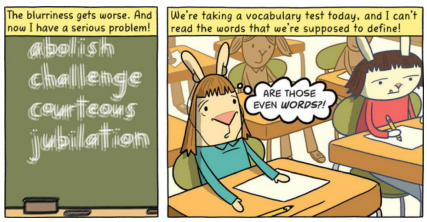
This paper has developed LFs and a semantic analyses for two varieties of free perception sequences in pictorial narratives, veridical ones and non-veridical ones. The semantics used situation calculus models where worlds are constructed as sequences of events. Perceptual information was modeled using event alternatives. A dynamic system of interpretation was used to take account of the fact that free perception is implicitly anaphoric.

While there is not space to talk about it, a goal in this enterprise is to develop connections and contrasts between phenomena in pictorial narratives, and analogous phenomena in linguistic narratives and current theoretical conceptions of them. Current work on free indirect discourse, such as Sharvit (2008), Eckardt (2015), and Hinterwimmer (to appear) is immediately relevant. This is mainly concerned with a broader category of free indirect discourse. But many of the data discussed by Hinterwimmer can be considered examples of linguistic free perception, see (19).[12] A good way to proceed here would be to analyze linguistic free perception using the formal tools that were used in this paper, and compare results.

(19) The T-Rex hesitated. Maybe the little dinosaurs had hidden themselves in the cave on his left. When Billy looked up in his hiding place a few seconds later, a T-Rex bent down to the entrance of the cave and squinted into the dark.

The handful of classes of examples discussed here do not exhaust the phenomena of pictorial free perception. We mention without comment a couple of cases that we conjecture require a different analysis. In one passage of Bell's *El Deafo*, Ceci has blurry vision. (21a) is a free perception sequence, showing her view of a blurred blackboard.[13] In the film *the Terminator*, the Terminator has an infrared visual system, and views from its perspective are rendered as in (21b), using a red palette.

(20) a.



b.

Perceptual phenomena can be rendered in bubbles. In a passage in *El Deafo*, Ceci has obtained a hearing aid, and has gained hyper-acuity to sound. In (21), she hears a teacher in the bathroom. Here there are issues of a disjuncture between auditory and visual information. While the bubble structure seems to indicate embedding, semantically Ceci's auditory

[12] At this writing, Hinterwimmer's work is available to us as a handout.
[13] The sequence is inverted, something that is possible also for veridical free-perception sequences.

information could not be strong engough to entail the visual information in the picture.

(21)

We could continue for quite a while the list of examples that should fall under an account of depiction of perception in pictorial narratives, but are not covered by what has been said here. We hope that what we have proposed is a good starting point.

References

Abusch, D. (2012). Applying discourse semantics and pragmatics to co-reference in picture sequences. In *Sinn und Bedeutung 17*.

Abusch, D. (to appear). Possible worlds semantics for pictures. In Lisa Matthewson, Cécile Meier, Hotze Rullmann, Thomas Ede Zimmermann (eds.), *Blackwell Companion to Semantics*.

Baltag, A., L. S. Moss, and S. Solecki (1998). The logic of public announcements, common knowledge, and private suspicions. In *Proceedings of the 7th conference on Theoretical Aspects of Rationality and Knowledge*, pp. 43–56. Morgan Kaufmann Publishers Inc.

Brinton, L. (1980). Represented perception: A study in narrative style. *Poetics 9*(4), 363–381.

Cumming, S., G. Greenberg, and R. Kelly (2017). Conventions of viewpoint coherence in film. *Philosopher's Imprint 17*(1).

Dekker, P. (2012). *Dynamic Semantics*. Springer.

Eckardt, R. (2015). *The Semantics of Free Indirect Discourse*. Brill.

Greenberg, G. J. (2011). *The Semiotic Spectrum*. PhD dissertation, Rutgers.

Hagen, M. A. (1980). *The Perception of Pictures. 1. Alberti's Window: the projective model of pictorial information*. Academic Press.

Hinterwimmer, S. (to appear). Two kinds of perspective taking in narrative texts. In *SALT 27*.

Kamp, H. and U. Reyle (1993). *From Discourse to Logic*. Springer.

Kuroda, S. (1994). Reflections on the foundations of narrative theory from a linguistic point of view. In *Noam Chomsky: From artificial intelligence to theology.(2 v.)*, pp. 779. Taylor & Francis.

Reiter, R. (2001). *Knowledge in Action: Logical foundations for specifying and implementing dynamical systems*. MIT Press.

Rooth, M. and D. Abusch (2017). Picture descriptions and centered content. In *Sinn und Bedeutung 21*.

Sharvit, Y. (2008). The puzzle of free indirect discourse. *Linguistics and Philosophy 31*(3), 353–395.

The scalar presupposition of *only* and *only if**

Sam Alxatib[1]

CUNY Graduate Center
alxatib@alum.mit.edu

Abstract

We diagnose a pattern of reversal in the scalar presupposition of *only* in *only if* constructions, and attempt to relate it to the monotonicity of conditional antecedents. At the heart of the analysis is a proposal that reduces the scalar presupposition of *only* to the particle's need to be non-vacuous. The reversal pattern is derived, but difficulties and questionable ingredients of the story are noted.

1 Introduction

This paper is about the scalar presupposition of *only* and its behavior in *only if* constructions. We focus our attention on the following generalization: while the use of *only* is dispreferred with relatively high focus associates, the same high associates are acceptable under *only if*, and it is with *low* associates that the acceptability of *only if* decreases. This is illustrated below:

(1) a. #This band only released ten_F albums (ten is high)
b. ✓This band only released two_F albums (two is relatively low)

(2) a. A band qualifies for this award only if they released (at least) ten_F albums
b. #A band qualifies for this award only if they released (at least) two_F albums

Let us make it clear from the start that we do not claim (1a) and (2b) to be categorically unacceptable. We merely highlight an apparent reversal in the effects of *only*'s scalar presupposition: being too high for acceptability with *only* coincides with being acceptable with *only if*, and being low and acceptable with *only* coincides with (near) unacceptability with *only if*. Note that the same reversal is found with alternative scales that are not logically ordered, as shown in (3-6) below. We will talk briefly about these cases later.

(3) a. ??John only got an A_F
b. ✓John only got a C_F

(4) a. A student will only be considered for admission if she gets (at least) an A_F
b. ??A student will only be considered for admission if she gets (at least) a C_F

(5) a. ??John only knows how to make $\text{turducken}_\text{F}$[1]
b. ✓John only knows how to make $[\text{boiled eggs}]_\text{F}$

(6) a. People get to work at that restaurant only if they know how to make $\text{turducken}_\text{F}$
b. ??People get to work at that restaurant only if they know how to make $[\text{boiled eggs}]_\text{F}$

*For helpful discussions, I thank Kai von Fintel, Danny Fox, Elena Herburger, Jon Nissenbaum, Yael Sharvit, and Anna Szabolcsi. All errors are my own.

[1] I'm assuming a scale of difficulty, and that turducken is hard to make, but boiled eggs are easy.

Similar findings seem to hold of bare plurals also. I will not discuss those in this paper.

(7) a. ??Only bands that released (at least) two$_{\text{F}}$ albums qualify for this award
b. ??Only students with (at least) a C$_{\text{F}}$ are considered for admission
c. ??Only people who know how to make [boiled eggs]$_{\text{F}}$ get to work at that restaurant

It is reasonable at first glance to relate this reversal to the downward monotonicity of *if*-clauses. Assuming that *only if* is composed of *only* and a conditional prejacent, and assuming that the focus associate of *only* in these cases is part of the antecedent, we expect the logical relationship between the alternatives to be reversed. This is because replacing an associate ϕ in 'if ϕ then ψ' with a stronger alternative ϕ' produces a *weaker* conditional. It would then follow that what counts as 'too strong' for association with *only* will make for a weak conditional prejacent in the case of *only if*, and that for a weak ϕ'', the conditional 'if ϕ'' then ψ' will be strong and thus (nearly) incompatible with *only*. This explains the reversal in (1-6).

The main goal of this paper is to lay out the details of this explanation. Doing this will involve making clear our assumptions about the semantics of *only if* constructions—here we will largely follow von Fintel 1997—and also involve articulating the scalar presupposition of *only* in a plausible way where the monotonicity of *if* will play this role. The formulation that I will suggest reduces the presupposition to another property that the particle is known to have: its infelicitousness when it is assertorically vacuous. The sketch of this reduction, and the predictions it brings to the *only*/*only if* reversal, is what I intend as the main contribution of the paper. To the extent that the overall proposal is plausible, a tentative corollary is that conditionals in *only if* constructions have universal (or near-universal) quantificational force. This contrasts with recent proposals in which *if* is assigned an existential semantics (Herburger 2015, Bassi and Bar-Lev 2017).

2 The semantics of *only* and *only if*

Standard analyses of *only* take the particle to operate on a propositional argument (the prejacent) and a set of alternatives to that argument. The alternatives are generated by replacing the focus-marked element in the prejacent with its contextually salient alternatives. Given a prejacent ϕ and a set A of alternatives to ϕ, *only* presupposes ϕ (though this is disputed)[2], and asserts the negation of whatever can be negated from among the elements of A. Consider (8):

(8) Mary only saw [John and Sue]$_{\text{F}}$

We analyze (8) effectively as an expression where *only* takes the sentence *John saw Mary* as its prejacent.[3] The alternatives in this case differ from the prejacent only with respect the focus-marked element *John and Sue*, giving us *Mary saw John*, *Mary saw Sue*, *Mary saw Bill*, etc. The semantics of *only*, shown in (9), negate those alternatives that do not follow from the prejacent, in this case, *Mary saw Bill*.

(9) Given a proposition ϕ and a set of propositions A,
$[\![only]\!]^w(A)(\phi)$ is defined only if $\phi(w)=1$, and if defined,
$[\![only]\!]^w(A)(\phi)=1$ iff $\forall\psi(\psi\in A \;\&\; \phi \not\models \psi \;\rightarrow\; \psi(w)=0)$

[2]The prejacent presupposition is due to Horn (1969). In Horn 1996 the presupposition is taken to be existential, and in Ippolito 2008 it is weakened further to a conditional presupposition. See Ippolito 2008 and Beaver and Clark 2008 for review and discussion of other possibilities.

[3]It is clear that *only* appears to take a VP argument here. We ignore this fact given that it does not affect the points of this paper.

Note that I set aside the mechanism with which the alternatives are made to depend on the form of the prejacent, and differ only in its focus. I refer the reader to Mats Rooth's work on this (Rooth 1985, Rooth 1992).

Let us extend the entry in (9) to *only if*, which we will take to consist of *only* with a conditional prejacent. As a working example, consider (10):

(10) Mary will only go if [John and Sue]$_{\text{F}}$ go

In (10), the focus associate of *only* appears inside the antecedent of the conditional prejacent. We do not want to say that every instance of *only if* is one where there is an identifiable focus-bearing expression inside the antecedent. But for now let us explore the possibility of analyzing (8) and (10) uniformly.[4]

Intuitively, (10) presupposes that Mary will go if John and Sue go (though again, this is not without controversy), and more relevantly for us, asserts that Mary will not go if John goes alone, and will not go if Sue goes alone. This reading is not straightforwardly derivable from the analysis developed so far. To see why, assume first a variably-strict implication account of the conditional prejacent, i.e. that *if* denotes a subsethood relation between accessible antecedent-worlds and consequent worlds:

(11) *If* as variably-strict
For any $p, q \in D_{\langle s,t\rangle}$, $[\![\textbf{if}]\!]^w(p)(q)=1$ iff $\text{SIM}_w(p) \subseteq q$
(where $\text{SIM}_w(p)$ is the set of maximally-similar p-worlds to w)

By our current assumptions, the alternatives to the conditional prejacent in (10) will look something like (12). Their negations, as provided by the assertion of *only*, are shown in (13).

(12) ALT(If [John and Sue]$_{\text{F}}$ go, Mary will go) = {If John goes, Mary will go,
If Sue goes, Mary will go, $\cdots$}

(13) $[\![(10)]\!]^w$ is defined only if $\text{SIM}_w(j\&s) \subseteq m$, and if defined
$[\![(10)]\!]^w=1$ iff $\text{SIM}_w(j) \not\subseteq m$ and $\text{SIM}_w(s) \not\subseteq m$ and $\cdots$

According to (13), the assertive component of (10) says that not all accessible (or maximally similar) John-going worlds are Mary-going worlds, and not all accessible Sue-going worlds are Mary-going worlds. But as von Fintel notes, this is not strong enough to capture the intuited meaning of (10). The conditions in (13) allow for some accessible John-going-alone worlds to be Mary-going worlds, so we predict that (10) be true in contexts where it is possible for Mary to go even if John goes without Sue. But intuitively, this is incorrect.

There are a number of ways of making the weak result above stronger. We will look at two of them, and we will point out an amendment that is needed on both. On the first option, we revise (11) and take *if* to denote an *existential* quantifier over worlds. This will do two things. It will weaken the truth conditions of conditionals generally, so we would then have to explain why they typically give rise to universal-like readings when unembedded.[5] But it will also provide us with a promising prediction: the negations of (existential) conditionals, the alternatives to the prejacent, will have strong truth conditions. The entry and its result are shown below.

[4] In Section 4 I will mention the possibility that, regardless of accenting, a conditional prejacent has only one alternative, that in which the antecedent is replaced with its negation. Unfortunately I will not be able to give this possibility the attention it deserves here.

[5] Bassi and Bar-Lev (2017) propose that the universal force of conditionals (in UE contexts) results from recursive exhaustification (Fox 2007).

(11′) An existential definition of *if*
For any $p, q \in D_{\langle s,t\rangle}$, $[\![\textbf{if}]\!]^w(p)(q)=1$ iff $\textsc{sim}_w(p) \cap q \neq \varnothing$
(where $\textsc{sim}_w(p)$ is the set of maximally-similar p-worlds to w)

(13′) $[\![(10)]\!]^w$ is defined only if $\textsc{sim}_w(j\&s) \cap m \neq \varnothing$, and if defined
$[\![(10)]\!]^w=1$ iff $\textsc{sim}_w(j) \cap m = \varnothing$ and $\textsc{sim}_w(s) \cap m = \varnothing$ and $\cdots$

The assertion in (13′) now says that no (maximally similar) John-going world is a Mary-going world, and no (maximally similar) Sue-going world is a Mary-going-world. However, we now have another problem. If some John-and-Sue-going worlds are Mary-going worlds, as the presupposition says, there is no way that *no* John-going worlds are Mary-going worlds, because the John-going worlds include John-and-Sue worlds, and we know that some of those are worlds where Mary goes. How do we get around this problem? Maybe we can assume that the maximally-similar worlds where John goes exclude those where he goes with Sue, but I am not prepared to discuss this possibility. Instead I will assume, at least given our current construal of the alternatives to conditionals, that the alternatives to the prejacent in *only if* constructions are conditionals whose antecedents are *exhaustified* with respect to the antecedent of the prejacent itself. In the case of the current example, this revision will give us (14).[6]

(14) ALT(If [John and Sue]$_F$ go, Mary will go) = {If $\textsc{exh}$(John goes), Mary will go,
If $\textsc{exh}$(Sue goes), Mary will go, $\cdots$}

With the revision in (14) we derive the desired assertion, as shown in (13″): the assertion says that no accessible John-but-not-Sue-going worlds are Mary-going worlds, and no accessible Sue-but-not-John-going worlds are Mary-going worlds.

(13″) $[\![(10)]\!]^w$ is defined only if $\textsc{sim}_w(j\&s) \cap m \neq \varnothing$, and if defined
$[\![(10)]\!]^w=1$ iff $\textsc{sim}_w(\textsc{exh}(j)) \cap m = \varnothing$ and $\textsc{sim}_w(\textsc{exh}(s)) \cap m = \varnothing$ and $\cdots$

Let us now turn to the second way of strengthening the weak results derived earlier. Here we will also need to maintain the internal-exhaustification assumption illustrated in (14), but instead of assuming an existential semantics for conditionals, we maintain universal force and add a homogeneity presupposition to them (von Fintel). We summarize this in (11″):

(11″) *If* as homogeneous and variably-strict
For any $p, q \in D_{\langle s,t\rangle}$, $[\![\textbf{if}]\!]^w(p)(q)$ is defined only if $\textsc{sim}_w(p) \subseteq q \vee \textsc{sim}_w(p) \subseteq \overline{q}$,
If defined, $[\![\textbf{if}]\!]^w(p)(q)=1$ iff $\textsc{sim}_w(p) \subseteq q$

According to (11″), conditionals impose an all-or-nothing precondition on their propositional inputs. When a conditional is false, it is false because the antecedent worlds are *disjoint* from the consequent worlds. This, together with the exhaustified alternatives in (14), produce a universal presupposition for *only if*, and also a strong assertion like the one in (13″):

(13″) $[\![(10)]\!]^w$ is defined only if $\textsc{sim}_w(j\&s) \subseteq m$, and if defined
$[\![(10)]\!]^w=1$ iff $\textsc{sim}_w(\textsc{exh}(j)) \subseteq \overline{m}$ and $\textsc{sim}_w(\textsc{exh}(s)) \subseteq \overline{m}$ and $\cdots$
i.e. iff $\textsc{sim}_w(\textsc{exh}(j)) \cap m = \varnothing$ and $\textsc{sim}_w(\textsc{exh}(s)) \cap m = \varnothing$ and $\cdots$

[6] This assumption is related to Menendez-Benito's (2005) Obligatory Exclusification Hypothesis, though I will leave a thorough comparison to a future occasion (I thank Kai von Fintel for pointing the similarity out to me).

Let us take stock. We followed von Fintel 1997 and assumed that *only if* constructions can be analyzed compositionally as cases where *only* takes a conditional prejacent. To make the analysis work, we revisited the important question of how to strengthen the exclusive component of *only if*. We then looked at two possible answers: on the first, we assume an existential semantics of *if*; on the second, we assume that conditionals carry a homogeneity presupposition. On either option we discovered that the antecedents in the alternative conditionals, assuming that they vary by the focus inside them, have to be understood to exclude the antecedent of the prejacent. We achieved this by stipulating that alternatives contain an embedded exhaustifier. The assumptions are summarized in (15), and the two options about the semantics of conditionals are shown in (15iii,iii′).

(15) (i) *Only if* consists of *only* together with a conditional prejacent.

(ii) The alternatives in the case of *only if* are conditionals that vary with respect to the focus associate in the prejacent, and they include conditionals where the antecedent is exhaustified against the antecedent of the prejacent.

(iii) Conditionals are variably-strict and homogeneous.

(iii′) Conditionals (under *only*) are existential.

3 The scalar presupposition of *only*

Everyone knows that *only* is evaluative. The intuition, illustrated earlier in (1,3,5), is sometimes captured by writing into the semantics of *only* a presupposition that its prejacent rank low with respect to its alternatives, on whatever ordering is provided in context (Klinedinst 2005, Zeevat 2008, Beaver and Clark 2008).

But what is the connection between the "height" of an alternative on a scale—the property that affects its acceptability as a prejacent to *only*—and the "height" of the conditional that contains that alternative in its antecedent? In what (possibly partial) way is the scale of conditionals based on the scale that its antecedent appears in, and what relationship is there between the threshold of lowness in one scale and the threshold of lowness in the other?

I will not attempt to answer these questions, because I want to try to reduce the scalar presupposition of *only* to another known constraint on the use of the particle. This is the ban against its assertoric vacuity, demonstrated below.

(16) a. #John only invited all_{F} of his friends

b. John only invited some_{F} of his friends

(17) a. #John only always_{F} puts sugar in his coffee

b. John only $\text{sometimes}_{\text{F}}$ puts sugar in his coffee

(18) a. #Of his three siblings, John only gets along with $[\text{Mary, Bill, and Sue}]_{\text{F}}$

b. Of his three siblings, John only gets along with $[\text{Bill and Sue}]_{\text{F}}$

The examples in (16-18) tell us that *only* is not licensed when it has no alternatives to negate — though for reasons that need not concern us, the more accurate characterization should say that *only* is infelicitous when its prejacent settles the truth values of all of its alternatives:

(19) $^{*}\textit{only}(p)$, given alternatives A, if $\forall p'(p' \in A \rightarrow (p \vDash p' \text{ or } p \vDash \neg p'))$

What determines the alternatives to a given prejacent? There are no doubt a number of formal constraints (see Katzir 2007 for a possible view), but beyond these, there must also be a number of contextual factors that allow some alternatives and not others to matter given the details of the conversational setting (see e.g. van Kuppevelt 1996). Notice for example that the acceptable (b) examples in (16-17) become strange with slight changes to the predicate:

(16b′) #John only stabbed some$_{\text{F}}$ of his friends

(17b′) #John only sometimes$_{\text{F}}$ puts sugar in his ears

As I said before, I do not claim these examples to be categorically infelicitous, but there is no denying that there are many imaginable natural contexts where they would sound odd or dismissible. Why should this be? There seems to be something beyond the formal and the scalar similarity of (16b,17b) to (16b′,17b′), and this may lead us to conclude that something additional to the vacuity ban takes part in the semantics of *only*. But I want to suggest that this conclusion is not necessary. It is also plausible that the oddness of (16b′,17b′) comes from a piece of common ground that makes the *some*/*sometimes* prejacents contextually-equivalent, respectively, to their *every*/*always* alternatives. These may be contexts where e.g. stabbing some friends and stabbing all of them are equally horrible, or where it is equally strange for John to sometimes put sugar in his ear as it is for him to always do so. If this is right, then the ban against vacuity *would* be violated in (16b′,17b′), because their prejacents happen to be contextually-equivalent to their formal universal alternatives, leaving nothing else for the exclusive particle to negate. The formal details of this idea, e.g. of how contextual equivalence can be represented and derived from the assumed conversational background, must be left for future work.[7]

Let us now assume an abstract set of alternatives $A=\{a_1, a_2, a_3\}$, and let a_3 asymmetrically entail a_2, and a_2 asymmetrically entail a_1:

(20) $a_1 \dashv a_2 \dashv a_3$

It is easy to see that within this group of alternatives, the ban against vacuity will make *only* infelicitous with a_3. This is because every alternative in A follows from a_3, and so *only* has no alternatives to negate, and is therefore assertorically vacuous. The cases of (16a,17a,18a) are instantiations of this case.

Consider now the case of *only if*, holding constant the assumptions in (15ii,iii), that *if* is variably-strict and homogeneous, and that its alternatives are determined by the alternatives to its antecedent. Here we predict vacuity in the case of [*only* [*if* a_1, q]], the weakest available antecedent, but not in the case of [*only* [*if* a_3, q]]. In the latter case the contribution of *only* will not be trivial because the assumed alternatives in (21) *are* predicted to be negated by the exclusive particle, as shown in (22). The assertive component of *only* will say that all worlds where a_1 is true but a_3 is false are worlds where $\neg q$, and likewise (redundantly) for worlds where a_2 is true but a_3 is false.

(21) $\text{ALT}(\textit{if}\ a_3,\ q) = \{\textit{if}\ \text{EXH}(a_1),\ q,$
$\qquad\qquad\qquad\qquad\qquad\qquad \textit{if}\ \text{EXH}(a_2),\ q\}$

[7] One possibility is to define "equivalence" as indistinguishability, and to base indistinguishability on plausible background considerations. Considerations can be represented as questions, which in turn are represented as sets of propositions. We now say that two alternatives (propositions) p, p' are indistinguishable relative to a question Q iff there is an answer q to Q such that both p, p' are subsets of q. This is intended to capture the intuition that p, p' do not provide different answers to Q, and are thus indistinguishable given Q.

(22) $[\![\textit{only}\ [\text{if}\ a_3,\ q]]\!]^w$ is defined only if $\text{SIM}_w(a_3) \subseteq q$, and if defined
$[\![\textit{only}\ [\text{if}\ a_3,\ q]]\!]^w = 1$ iff $\text{SIM}_w(\text{EXH}(a_2)) \subseteq \overline{q}$ and $\text{SIM}_w(\text{EXH}(a_1)) \subseteq \overline{q}$ and $\cdots$

But what if the weakest alternative a_1 appears in the antecedent of *only if*? In this case we predict an infelicitous use of *only*, on account of vacuity. The alternative set is shown in (23):

(23) $\text{ALT}(\textit{if}\ a_1,\ q) = \{\textit{if}\ \text{EXH}(a_2),\ q,$
$\textit{if}\ \text{EXH}(a_3),\ q\}$

In each alternative in (23) the antecedent entails the antecedent of the prejacent.[8] Therefore, on the strict implication view the alternatives come out to be weaker than the prejacent, so they are not negated by *only*. The overall result, then, is that given a set of logically-ordered alternatives like (20), *only* is predicted to be vacuous with the strongest element, and in the case of *only if* the vacuity is predicted if the antecedent contains the weakest element. This holds if we assume (15ii,iii): strict-implication and associate-driven alternatives.

Can we find a vacuous *only if* that instantiates this case? As a first example suppose we take the *some-all* scale. If we can be sure that the scale is limited to just these two items, or at least that it contains nothing weaker than *some*, then we predict that *only if* containing a *some*-antecedent be infelicitous, but this isn't true:

(24) Mary will only go if some_F of her friends go

But perhaps the conditional here has an alternative where *some* is replaced by *no*. If so, then we no longer predict vacuity.[9] Another kind of example we might look for is one where the antecedent is trivially weak. (25) is an example, and it is indeed strange.

(25) #John will only buy the car if it has (at least) two doors

But the construction is also strange without *only*:

(26) #John will buy the car if it has (at least) two doors

The trouble here is that the trivial antecedent makes the conditional equivalent to its consequent. This alone may be why both (25) and (26) are odd. We may therefore be up against a design confound: the kind of conditional that would instantiate $[\textit{if}\ a_1,\ q]$ may be the very same kind of conditional that is equivalent to its consequent, and hence infelicitous independently. What we need is a case of a licit conditional where the antecedent is for all intents and purposes vacuous, but which is still used acceptably to communicate its consequent. (27a,b) are examples of this sort, and indeed, they are quite strange in their *only if* versions:

(27) a. If the car gets him from A to B, he will buy it
b. If he wakes up breathing, he will go to his daughter's wedding

(28) a. #He will only buy the car if it gets him from A to B
b. #He will only go to his daughter's wedding if he wakes up breathing

[8] This is true regardless of the contribution of EXH; because a_2 and a_3 are by assumption stronger than a_1, and $\text{EXH}(a_2)/\text{EXH}(a_3)$ are either stronger or equivalent to a_2/a_3, it follows that the antecedents of the alternatives in (23) entail a_1.

[9] I think there are independent empirical reasons to keep *no* out of the *some-every* scale, but I can't discuss them here. Matsumoto (1995) has argued that formal alternatives should have the same monotonicity, and if he is right then we cannot use *no* to rescue (24).

This is as much as I can do to find a convincing instance of a vacuous, and hence infelicitous, *only if*. Now I want to relate the discussion to the scalar presupposition of *only*.

Take a scale where some background information makes alternatives contextually equivalent. An example is the case of *sometimes put sugar in one's ear* and *always put sugar in one's ear*. Assuming that doing either is equally weird, and assuming that the conversational background does not concern finding finer grades of weird behavior, the distinction between *some* and *every* in this case will be blurred, and this causes the alternatives to occupy the same node in the scale. From this perspective, we expect adjacent nodes within a given scale to be more susceptible to collapse than non-adjacent nodes. We also expect vacuity of *only* to be more likely when its prejacent is high than when it is low; with a high prejacent, equivalence to nearby higher alternatives brings the prejacent closer to the end of the scale, thus closer to making *only* vacuous. This is not true of lower prejacents. However, we expect the reverse for *only if*. Presumably, if a_i and a_j are contextually equivalent, then the conditionals $[\textit{if}\ a_i,\ q]$ and $[\textit{if}\ a_j,\ q]$ will also be contextually equivalent. An instance of *only if* that contains a low antecedent has a greater chance of being vacuous than one that contains a high antecedent.

Let me summarize. I have suggested that what researchers call the scalar presupposition of *only* is the same as the particle's need to be assertively non-vacuous. The inference arises in its guise as a separate presupposition in just those cases where the only alternatives that can be negated happen to be in some sense contextually-equivalent to the prejacent. This keeps them from being excluded by the particle, and the particle is consequently made vacuous. Assuming this perspective, we saw that the higher elements of a scale of alternatives are more likely to give rise to these near-vacuity violations under *only*, and that the lower ones are the more likely to cause near-vacuity for *only if*. This was the reversal that we wanted to capture.

4 Remaining issues and concluding remarks

The sketch presented in this paper makes many theoretical presumptions. Among them is that the alternatives to *if* in *only if* are determined by changing the associate in the *if*-clause with its scalemates. Another plausible take on this is that conditional prejacents have only one alternative: that in which the antecedent is replaced with its negation. I have not addressed this possibility in this paper for reasons of space, and I leave it for future work. An important question is whether *only if* can ever be vacuous if the alternative to the prejacent $[\textit{if}\ \phi,\ \psi]$ is the conditional $[\textit{if}\ \neg\phi,\ \psi]$. Vacuity here would require the two conditionals to be equivalent in some contextually determined sense, but I do not yet know how this might work in a principled way. If it cannot work, and if there are good reasons to adopt this stance on alternatives, then what I proposed is likely wrong.

On the other hand, if this proposal is on the right track, it sheds light on a couple of issues. One of them concerns the quantificational force of *if* under *only*. We saw earlier that, on the variably-strict treatment, *only if* is predicted to be vacuous when its antecedent is the weakest in the given scale. But this prediction does not follow if *if* is existential (recall (15iii′)). To see why, take our abstract scale again:

(29) $a_1 \dashv a_2 \dashv a_3$ (=(20))

If the prejacent contains the weakest member of the scale, as in $[\textit{if}\ a_1,\ q]$, then we have the alternatives in (30).

(30) $\text{ALT}(\textit{if}\ a_1,\ q) = \{\textit{if}\ \textsc{exh}(a_2),\ q,\ \textit{if}\ \textsc{exh}(a_3),\ q\}$ (=(23))

But on an existential view, the alternatives are stronger than the prejacent, because they make existential claims about a smaller set of worlds than the prejacent does. In this case [*only* [*if* a_1, q]] should mean that some a_1 worlds are q worlds, and that no a_2 worlds are q worlds, and no a_3 worlds are q worlds. The relationship between the scale and the position in it that leads to vacuity will not emerge in the way it did on the strict-implication view. Again, however, I must reiterate that the validity of this point rests on our assumption (15ii) about alternatives.[10]

Finally, I have only discussed scales in which alternatives are ordered by their logical strength. But as I noted, reversal holds also in cases where the alternatives are non-logically ordered (recall (3-6)). If the vacuity account of reversal is right, along with our other assumptions about alternatives and the semantics of *if*, then the findings suggest that *only* is logical even when the contextually understood alternatives are ordered non-logically. In those cases, *only* operates on a reinterpretation of the contextually provided ranking, where each element corresponds to the disjunction that consists of it and every scalemate above it. This way, the scalar ordering is translated to a logical ordering, and given the logical ordering, the predictions derived above would hold in the same way. The details of this must be left for future development.

References

Bassi, Itai, and Moshe Bar-Lev. 2017. A unified existential semantics for bare conditionals. In *Sinn und Bedeutung 21*, ed. Rob Truswell.

Beaver, David, and Brady Clark. 2008. *Sense and Sensitivity*. Wiley Blackwell.

von Fintel, Kai. 1997. Bare plurals, bare conditionals, and *only*. *Journal of Semantics* 14:1–56.

Fox, Danny. 2007. Free choice and the theory of scalar implicatures. In *Presupposition and Implicature in Compositional Semantics*, ed. Uli Sauerland and Penka Stateva. Houndmills: Palgrave Macmillan.

Herburger, Elena. 2015. *Only if*: if only we understood it. In *Sinn und Bedeutung 19*, ed. Eva Csipak and Hedde Zeijlstra.

Horn, Laurence R. 1969. A presuppositional analysis of *only* and *even*. In *CLS 5*, ed. Robert I. Binnick, Alice Davidson, Georgia M. Green, and Jerry L. Morgan. University of Chicago Department of Linguistics.

Horn, Laurence R. 1996. Exclusive company: *only* and the dynamics of vertical inference. *Journal of Semantics* 13:1–40.

Ippolito, Michela. 2008. On the meaning of *only*. *Journal of Semantics* 25:45–91.

Katzir, Roni. 2007. Structurally-defined alternatives. *Linguistics and Philosophy* 30:669–690.

[10] Bassi and Bar-Lev propose an existential semantics of conditionals, but add subdomain alternatives. Though each subdomain alternative would on their view be stronger than the conditional prejacent, together the subdomain alternatives exhaust the worlds that make up the antecedent. This makes the alternatives non-innocently excludable. So if all these subdomain alternatives are added to the stronger alternatives entertained in this paper, the predictions will change and will make it possible to derive similar vacuity predictions. However, questions still remain about alternative conditionals with weaker antecedents. Under strict implication these are stronger and hence potentially excludable, but on an existential semantics they are weaker globally and hence unexcludable.

Klinedinst, Nathan. 2005. Scales and *Only*. Master's thesis, UCLA.

van Kuppevelt, Jan. 1996. Inferring from topics: Scalar implicatures as topic-dependent inferences. *Linguistics and Philosophy* 19:393–443.

Matsumoto, Yo. 1995. The conversational condition on Horn Scales. *Linguistics and Philosophy* 18:21–60.

Menendez-Benito, Paula. 2005. The Grammar of Choice. Doctoral Dissertation, UMass Amherst.

Rooth, Mats. 1985. Association with focus. Doctoral Dissertation, UMass Amherst.

Rooth, Mats. 1992. A theory of focus interpretation. *Natural Language Semantics* 1:75–116.

Zeevat, Henk. 2008. "Only" as a mirative particle. In *Focus at the Syntax-Semantics Interface*, ed. Arndt Riester and Edgar Onea. Working Papers of the SFB 732, Vol. 3, University of Stuttgart.

Dislocated Cosuppositions

Amir Anvari*

Institut Jean Nicod, Département d'Études Cognitives,
École Normale Supérieure, PSL Research University,
CNRS, Paris, France
amiraanvari@gmail.com

1 Introduction

As a preliminary illustration of the problem this paper is concerned with, consider the sentence in (1). [*On notation:* a speech accompanying (or, co-speech) gesture is notated as a subscript in SMALL CAPITALS after the expression it co-occurs with. The modified expression is put between square brackets if it contains several words.]

(1) a. John punished$_{\text{SLAP}}$ his son.[1]
   ⇝ John punished his son by slapping him

   b. John [took the elevator]$_{\text{UP}}$.[2]
   ⇝ John took the elevator to go up

In each case, the co-occurring gesture enriches the basic meaning of the sentence in a manner that is clearly keyed to its iconic shape. As with any other form of enrichment, one may ask three major questions about gestural enrichments: (i) what is the *form* of the gestural enrichments?,[3] (ii) what is the *projection profile* of gestural enrichments?, and (iii) what is the *epistemic status* of gestural enrichments?. Building on [3], this paper aims at contributing to each of these questions. Beginning with question (ii), made more explicit in (2), we need to embed gesturally modified expressions in the scope of logical operators and inquire about the fate of the gestural inference as it projects through these operators. The salient case of negation is given in (3).

(2) **The projection problem for co-speech gestures.** How are the enrichments of expressions modified by co-speech gestures inherited by complex sentences? (from [3], see also the pioneering work of [1])

(3) a. John did not punish$_{\text{SLAP}}$ his son.
   ⇝ if John had punished his son, he would have done so by slapping

   b. John did not [take the elevator]$_{\text{UP}}$.
   ⇝ if John had taken the elevator, he would have done so to go up

*I am greatly indebted to Philippe Schlenker and Benjamin Spector. All errors are emphatically mine.

[1] 'SLAP' stands for a slapping gesture in "neutral position" (i.e. close to torso).

[2] 'UP' stands for an upward movement of arms.

[3] Example: the form of the scalar implicature associated with a sentence of the form 'some As B' is 'not all As B', the form of the homogeneity inference associated with a sentence of the form 'the As B' might be taken to be 'either all As B or all As not B', etc.

The judgments reported in (3) (following '⇝') become sharper once appropriate context is provided. For example, compare an utterance of (3b) "out of the blue" versus in the context specified in (4).

(4) [Context: the building has ten floors. Mary's office is on the 5th. We do not know where John's office is. John does not know where Mary's office is. He has been looking for her.]
A: Did John manage to find Mary's office?
B: No... he got lost on the 5th floor...
A: How did *that* happen? Her office is right in front of the elevator!
B: Well, he didn't [take the elevator]$_{\text{UP}}$, he used the stairs instead.

The inference suggested in (3b) is quite sharply felt in (4): John's office (or at least his starting point before he went looking for Mary) is on a floor below the 5th: he did not use the elevator, but *if he had done so, he would have gone up.*

2 The Cosuppositional Analysis

The starting point of this paper is [3]'s "cosuppositional" analysis of gestural enrichments. This analysis takes the form of the judgments provided in (3) quite seriously. With a good deal of simplification, it can be summarized as follows.

(5) The Cosuppositional Approach. (hf. CA) If a predicate α embedded in a sentence ϕ uttered in context C is accompanied by a gesture G, the local context of α in ϕ relative to context C must entail $\boldsymbol{\alpha} \Rightarrow \boldsymbol{G}$:[4]

$$\vDash_{lc(\alpha)} \boldsymbol{\alpha} \Rightarrow \boldsymbol{G}.$$

In words, the cosuppositional analysis requires that the gesturally modified expression must entail the content of the accompanying gesture in its local context. The intuition behind this requirement is that iconic, co-verbal gestures *illustrate* the *local* meaning of the expressions they modify. "Local meaning" is here understood as semantic denotation relative to a given local context, where the latter is formulated on the basis of [2]'s theory of local contexts. Thus, if α is some predicate and $lc(\alpha)$ is its local context, the local meaning of α boils down to $lc(\alpha) \wedge \boldsymbol{\alpha}$. "Illustration" is quite simply cashed out as entailment. The requirement, therefore, can be formalized as $\vDash_{lc(\alpha)\wedge\boldsymbol{\alpha}} \boldsymbol{G}$ which is equivalent with $\vDash_{lc(\alpha)} \boldsymbol{\alpha} \Rightarrow \boldsymbol{G}$. The analysis, thus, is tantamount to saying that a predicate/gesture complex 'α_{G}' triggers the presupposition that $\boldsymbol{\alpha} \Rightarrow \boldsymbol{G}$.

Schlenker's CA answers the three questions posed at the beginning of this paper as follows: (i) gestural enrichments are pieces of information that are conditionalized on the assertive content of the expressions they modify, (ii) gestural enrichments project like presuppositions do in general, and (iii) gestural enrichments receive the same epistemic treatment as root as presuppositions, namely they must be entailed by the Common Ground[5] (for the utterance to be acceptable).

[4] Here and throughout: for any expression α, $\boldsymbol{\alpha} = [\![\alpha]\!]$. For a gesture G, $\boldsymbol{G}$ is also taken to be the model-theoretic object it "denotes".

[5] Common Ground: the conjunction of all propositions that the interlocutors take for granted at a particular point of a conversation. Context Set: the set of all possible worlds that are compatible with the Common Ground. See Stalnaker.

CA accounts for the judgments reported in (3) immediately: presuppositions project from under negation, therefore, e.g., (3a) is predicted to put the following requirement on the Context Set, C: any world w in C is such that either John did not punished his son in w, or John punished his son by slapping in w (i.e., C entails that John did not punished his son by any mean other than slapping). Exactly the same prediction is made for the unembedded case, (1b). The predicted "net effect" is of course correct: if C entails that John did not punished his son without slapping him, adding the information that John *did* punish his son will contextually convey that John punished his son by slapping him.[6]

CA also makes welcome predictions for the cases of embedding gesturally modified expressions in the scope of the quantifiers 'every' and 'no', (6). As is well-established, presuppositions project universally out of the scope of 'every' and 'no'. The cosupposition associated to the predicate '$\lambda x.$ x punished$_{\text{SLAP}}$ x's son' is the property [$\lambda x.$ **punished**(x,x's son) $\Rightarrow$ **slapped**(x,x's son)]. Once this presupposition is projected universally to root, one gets the predicted inferences in (6) which line up nicely with the attested inferences.

(6) a. Each of these ten guys punished$_{\text{SLAP}}$ his son.
$\rightsquigarrow$ Each of the guys punished his son by slapping him (attested)
$\rightsquigarrow$ $\forall x \in \textbf{\textit{guys}} : \textbf{\textit{punished}}(x, x\text{'s son}) \Rightarrow \textbf{\textit{slapped}}(x, x\text{'s son})$ (predicted)

b. None of these ten guys punished$_{\text{SLAP}}$ his son.
$\rightsquigarrow$ Each of the guys would have slapped his son, had he punished him (attested)
$\rightsquigarrow$ $\forall x \in \textbf{\textit{guys}} : \textbf{\textit{punished}}(x, x\text{'s son}) \Rightarrow \textbf{\textit{slapped}}(x, x\text{'s son})$ (predicted)

However, as Schlenker points out, the predictions made by CA are in some cases *too strong*. This is in particular the case for non-monotonic environments.

(7) a. Mary is unaware that John punished$_{\text{SLAP}}$ his son.
$\rightsquigarrow$ John punished his son by slapping him (attested)
$\rightsquigarrow$ $\textbf{\textit{punished}} \wedge (\textbf{\textit{punished}} \Rightarrow \textbf{\textit{slapped}}) \wedge \text{B}_{\text{M.}}(\textbf{\textit{punished}} \Rightarrow \textbf{\textit{slapped}})$[7] (predicted)

b. Some but not all of these ten guys punished$_{\text{SLAP}}$ their son.
$\rightsquigarrow$ Some of the guys punished their son by slapping, the rest did not punished their sons in any way (attested)
$\rightsquigarrow$ $\forall x \in \textbf{\textit{guys}} : \textbf{\textit{punished}}(x, x\text{'s son}) \Rightarrow \textbf{\textit{slapped}}(x, x\text{'s son})$ (predicted)

Consider (7a). It is reasonable to analyze a sentence of the form 'S is unaware that P' as presupposing that P and asserting that it is not the case that S believes that P, $\neg \text{B}_{\text{S}}$. Therefore, regarding presuppositions triggered in the subordinate clause, we predict that, first, these must project to root (8a) and, second, these must be entailed by the beliefs of the attitude holder (8b).

(8) Mary is unaware that John has stopped smoking
a. $\rightsquigarrow$ John used to smoke but no longer does
b. $\rightsquigarrow$ Mary believes that John used to smoke

[6] Just why the conditional force of the inference is not felt for the unembedded cases in (1) is a question that I will follow Schlenker by ignoring.

[7] Here '***punished***' is short for '***punished***(***John***, ***John's son***)'. Same with '***slapped***'. For any P, '$\text{B}_{\text{M.}}$(P)' stands for 'Mary believes that P'.

The problem raised by (7a) is that an utterance of (7a) can easily be understood such that only the first of these prediction is born out. The sentence itself presupposes that John punished his son, since the cosupposition that if John punished, he slapped also projects to root, we predict the overall presupposition that John punished his son by slapping him. But, the second prediction (namely, that Mary believes that John did not punish his son without slapping him), if available at all, is not easily accessible.

The problem raised by (7b) is similar: presuppositions triggered in the scope of the complex determiner 'some but not all' project universally to root, (9); consequently, CA predicts, not only that some guys punished their son by slapping, but also that for each of the guys who did not punished their son, if they had done so, they would have slapped. This latter inference is at least not easily accessible (but see the discussion in section 4); (7b) can very naturally be understood to imply that those guys who did in fact punish their son did so by slapping, without making any implication about the punishing habits of the other guys.

(9) Some but not all students have stopped smoking.
⇝ Every student used to smoke

In the next section I will discuss a solution to the problems raised in (7) which is formulated by Schlenker himself. Once the limits of that solution are made explicit, I will turn to my own proposal in section 4.

3 The "Supervaluationist" theory

Let us go back to the problem raised by 'unaware' in (7a) repeated below.

(10) Mary is unaware that John punished$_{\text{SLAP}}$ his son.
⇝ John punished his son by slapping him (attested)
⇝ ***punished*** $\wedge$ (***punished*** $\Rightarrow$ ***slapped***) $\wedge$ B_{M}(***punished*** $\Rightarrow$ ***slapped***) (predicted by CA)

Consider the following line of attack. What happens when a gesture modifies an expression, as in (10), is that two propositions are made salient for the audience to choose from. In the case of (10) these could be (11a) and (11b).

(11) a. That Mary is unaware that John punished his son.
$\underline{\text{P}} \wedge \neg\text{B}_{\text{M}}(\text{P})$[8]

b. That Mary is unaware that John punished his son by slapping him.
$\underline{(\text{P} \wedge \text{S})} \wedge \neg\text{B}_{\text{M}}(\text{P} \wedge \text{S})$

What would the audience do, when they are faced with such a choice? One possible answer is that the audience are inherently conservative: they "focus attention" only to those situations in which both propositions in (11) are simultaneously true (/false). In other words, they assume the speaker would not make an utterance like (10) if he believes that the two propositions in (11) have distinct truth-values. The prediction, then, is that an utterance of (10) is true (/false) iff both propositions (11a) and (11b) are true (/false). Interestingly, this prediction is *weaker* than the one made by CA. Since $\text{P} \wedge \text{S}$ is stronger than P while $\neg\text{B}_{\text{M}}(\text{P})$ is stronger than $\neg\text{B}_{\text{M}}(\text{P} \wedge \text{S})$, (10) is predicted to be true if and only if $(\text{P} \wedge \text{S}) \wedge \neg\text{B}_{\text{M}}(\text{P})$. No problematic inference is predicted pertaining to Mary's beliefs, as desired.

[8] Underlining marks for presuppositionality.

The general principle underlying the reasoning spelled out in the previous paragraph can be summarized as follows.

(12) The "Supervaluationist" Analysis.[9] (hf. SA) Let ϕ be a sentence that contains the predicate α, $\phi = \phi[\alpha]$. An utterance of $\phi[\alpha_{\mathrm{G}}]$ is judged true (false) iff both $\phi[\alpha]$ and $\phi[\alpha \wedge \mathrm{G}]$ are true (resp. false).

Here is another example that is adequately dealt with by SA.

(13) Exactly one of these ten guys punished$_{\textsc{slap}}$ his son.
$\rightsquigarrow$ Exactly one of the guys punished his son by slapping, the rest did not punished their sons in any way

Since CA is built on the Transparency Theory as its projection engine,[10] it predicts the cosupposition triggered in the scope of 'exactly one' in (13) to project universally to root, quite the same as the case of 'some but not all'. The result is the correct prediction that one guy punished his son by slapping and the rest did not punish their son and the *in*correct prediction that for each of the guys who did not punish their son, if they had done so, they would have slapped. Here again, the prediction made by SA is adequately weak; as the reader can easily verify, if an utterance of (13) is true iff both (14a) and (14b) is true, then an utterance of (13) is true iff one guy punished his son by slapping and the rest did not punish their son in any way. No inference is predicted regarding the guys who did not punish their son, as desired.

(14) a. Exactly one of these ten guys punished his son.
b. Exactly one of these ten guys punished his son by slapping him.

Unfortunately, SA has problems of its own (which Schlenker points out). Specifically, the predictions made by SA are sometimes *too weak*, sometimes to the point of triviality. For example, the prediction made for (7b), repeated below, is that it is true iff some guys punished their son by slapping and some guys did not punish their son in any way; this is too weak, as it allows for there being guys who punished their son in some way other than by slapping.

(15) Some but not all of these ten guys punished$_{\textsc{slap}}$ their son.
a. Some but not all of these ten guys punished their son.
b. Some but not all of these ten guys punished their son by slapping him.

Further, when a gesturally modified expression is embedded in a Downward Entailing environment, SA predicts *no enrichment* to the truth-conditions of the the sentence. For example, (6b), repeated below, is predicted to be true iff none of the guys punished their son in any way. The reason being that since (16a) entails (16b), the requirement that both be true boils down to the requirement that (16a) be true.

[9] This principle is *reminiscent* of the type of reasoning that supervaluationist logics are known for, hence the title and the quotation marks.
[10] Transparency Theory predicts in general presuppositions triggered in the scope of quantifiers projects universally to root.

(16) None of these ten guys punished$_{\text{SLAP}}$ his son.

a. None of these ten guys punished his son.

b. None of these ten guys punished his son by slapping him.

To recap (and repeat), the predictions made by CA are sometimes too strong while those made by SA are sometimes too weak. One might wonder whether the two should be put together. There are two main obstacles to this idea. First, SA and CA seem two entirely distinct mechanisms, a marriage between the two (regardless of the exact details) seems hopelessly disjunctive ("conceptually odd" in Schlenker's words). Second, it is not entirely clear just how the two analyses must be "linked" together. To see this, consider Schlenker's own suggestion.

(17) A co-speech gesture is treated in terms of SA (= (12)) unless this fails to strengthen the meaning, in which case it is treated in terms of CA (= (5)).

This way of linking CA and SA immediately runs into a problem with (15): in that case, as I have noted, SA *does* strengthen the meaning, but it does not do so sufficiently. In the next section I will formulate a proposal that solves these two problems (i.e., the linking problem and the problem of conceptual oddity) in one stroke. I will then show that this new principle coupled with a new bridge principle to link the predicted inferences with the background assumptions yields empirically adequate predictions.

4 Dislocated Cosuppositions

To spell out my proposal, I need to define several auxiliary notions. Let α be a predicate, and ϕ a sentence that contains (an occurrence of) α. We can construct a sequence β_i of property- or proposition-denoting constituents of ϕ with the following properties: (i) $\beta_0 = \alpha$, (ii) $\beta_n = \phi$, and (iii) for each $i \in \{0, \ldots, n-1\}$, $\beta_i \sqsubseteq \beta_{i+1}$ (β_i is contained in β_{i+1}). Let me call this the formation sequence of ϕ relative to α. Further, given a Context Set C, we can annotate each β_i with its local context, $lc(\beta_i)$, given [2]'s algorithm.[11] Finally, I need a notion of logical/contextual entailment which applies to property- and proposition-denoting expressions.

(18) Let β and β' be two expressions of a type that 'ends in t' which can take n arguments. let C (the "context") be a model-theoretic object of the same type. Then,

a. $\beta \vDash \beta'$ iff for all objects $x_1, \ldots, x_n$ of appropriate types, if $[\![\beta]\!](x_1)\ldots(x_n) = 1$, then $[\![\beta']\!](x_1)\ldots(x_n) = 1$.

b. $\beta \vDash_{\text{C}} \beta'$ iff for all objects $x_1, \ldots, x_n$ of appropriate types, if $\text{C}(x_1)\ldots(x_n) = 1$ and $[\![\beta]\!](x_1)\ldots(x_n) = 1$, then $[\![\beta']\!](x_1)\ldots(x_n) = 1$.

My proposal can now be formulated as follows.

(19) The Dislocated-Cosuppositions Analysis. (hf. DC) Let ϕ be a sentence that contains the predicate α, and let $\langle \beta_0 = \alpha, \ldots, \beta_n = \phi \rangle$ be the formation sequence of α relative to ϕ, and let G be some gesture. An utterance of $\phi[\alpha_{\text{G}}]$ is admitted by a context C only if there is some $i \in \{0, \ldots, n\}$ such that (i) $\beta_i[\alpha] \nvDash \beta_i[\alpha \wedge \text{G}]$ but (ii) $\beta_i[\alpha] \vDash_{lc(\beta_i)} \beta_i[\alpha \wedge \text{G}]$. If felicitous in C, $\phi[\alpha_{\text{G}}]$ is interpreted as $\phi[\alpha]$.

[11] The more accurate notation is $lc(\text{C}, \beta_i, \phi[\,\cdot\,])$.

The reasoning that is compressed in (19) can be unpacked as follows. Consider an utterance of $\phi[\alpha_{\text{G}}]$, where α is a predicate and G is a co-occurring gesture. For each constituent β of ϕ that contains α, a "gestural alternative" can be constructed by conjoining the "meaning" of G with α, $\beta[\alpha \wedge \text{G}]$.[12] Among these constituents, one can identify those that do not semantically entail their gestural alternatives. Then, the utterance is acceptable in C as soon as one of these constituents *contextually* entails its gestural alternative (in its local context).

I would like to make three remarks immediately. First, it is always the case that the inference generated by $\beta_0 = \alpha$ is identical with the cosupposition predicted by CA. Second, for every example discussed in this paper, the inference generated by $\beta_n = \phi$ is identical with the one generated by SA. This, indeed, in the sense in which CA and SA can be viewed as the outcomes of the same algorithm applied locally and globally. Third, (19) as it stands predicts "intermediate" inferences. I have not been able to construct good examples to establish whether this is a good or a bad prediction, but it should be clear that in case there are no such intermediate inferences (19) can be reformulated to make reference only to the most global and the most local constituents. This issue will not be relevant in the rest of this paper.

I will now work through the examples discussed above to evaluate the predictions of (19). Let us begin with the case of the universal quantifier 'every'.

(20) Every one of these ten guys punished$_{\text{SLAP}}$ his son.

a. $\beta_0 = \lambda x.\ x$ punished x's son
$lc(\beta_0) = \lambda w.\lambda x.\ w \in \text{C} \wedge x$ is one of the guys in w

b. $\beta_1 =$ [every guy] [$\lambda x.\ x$ punished x's son]
$lc(\beta_1) = \lambda w.\ w \in \text{C}$

In the case of (20) since the local context of the scope (viewed extensionally) is simply the set of all guys, the inference triggered by both (20a) and (20b) boils down to the same; (20a) predicts the inference that for each guy g, if g punished his son, he slapped him and (20b) predicts the inference that if every guy punished his son, then every guy punished his son by slapping. This is of course the same prediction that CA makes, which in conjunction with what the sentence (20) (without the gesture) asserts, yields the attested inference that every guy punished his son by slapping him. Next, consider the case of the negative quantifier 'no' (which, remember, was problematic for SA).

(21) None of these ten guys punished$_{\text{SLAP}}$ his son.

a. $\beta_0 = \lambda x.\ x$ punished x's son
$lc(\beta_0) = \lambda w.\lambda x.\ w \in \text{C} \wedge x$ is one of the guys in w

b. $\beta_1 =$ [no guy] [$\lambda x.\ x$ punished x's son]
$lc(\beta_1) = \lambda w.\ w \in \text{C}$

Here, no inference is predicted to arise by (21b) because β_1 logically entails $\beta_1[\alpha \wedge \text{G}]$ (= [no guy] [$\lambda x.\ x$ punished x's son by slapping]), violating the condition (i) of (19). The only option, therefore, is for (21a) to trigger an inference, which, as with (20a), boils down to the presupposition that for each guy g, if g punished his son, he slapped him. This is again the same (correct) prediction that CA makes.

Let me know move on to the case of 'unaware' (which was problematic for CA).

[12] I am, of course, conflating meta- and object-languages here. This is merely to avoid clutter.

(22) Mary is unaware that John punished$_{\text{SLAP}}$ his son.

a. $\beta_0 =$ John punished his son
$lc(\beta_0) = \lambda w.\lambda w'.\ w \in \text{C} \wedge w' \in (\text{DOX}^{w}_{\text{M.}} \cup \{w\})$[13]

b. $\beta_1 =$ Mary is unaware that John punished his son
$lc(\beta_1) = \lambda w.\ w \in \text{C}$

DC predicts two possible inferences for (22). One option is (22a), which will generate the same the prediction as the one made by CA. The second option is (22b), which will generate the same the prediction as the one made by SA. Before elaborating on this ambiguity, let me also mentioned another example, involving 'exactly one'.

(23) Exactly one of these ten guys punished$_{\text{SLAP}}$ his son.

a. $\beta_0 = \lambda x.\ x$ punished x's son
$lc(\beta_0) = \lambda w.\lambda x.\ w \in \text{C} \wedge x$ is one of the guys in w

b. $\beta_1 =$ [exactly one guy] [$\lambda x.\ x$ punished x's son]
$lc(\beta_1) = \lambda w.\ w \in \text{C}$

Here again, the inference predicted by (23a) is the same as CA, while it can easily be verified that the inference predicted by (23b) is that of SA. Now, is the ambiguity predicted by DC regarding, e.g., (22) and (23) undesirable? Not necessarily. Although the facts are at the moment rather unclear, ? find that 'exactly one' at least sometimes gives rise to universal inferences. The important point, for my purposes was to construct a system which can derive the inferences that Schlenker's CA could not. But the resulting system predicts systematic ambiguity. The evaluation of this prediction needs to be postpones until the facts are cleared up.

Finally, let me point out that one problem still remains, having to do with 'some but not all' (the same point can be made with 'between n and m', 'an odd number of', etc.).

(24) Some but not all of these ten guys punished$_{\text{SLAP}}$ his son.

a. $\beta_0 = \lambda x.\ x$ punished x's son
$lc(\beta_0) = \lambda w.\lambda x.\ w \in \text{C} \wedge x$ is one of the guys in w

b. $\beta_1 =$ [some but not all guy] [$\lambda x.\ x$ punished x's son]
$lc(\beta_1) = \lambda w.\ w \in \text{C}$

The problem is that since the predictions made by DC match those made by CA and SA, DC *cannot* account for (24); the prediction made on the basis of (24a) is too strong while the one made on the basis of (24b) is too weak. This is indeed the same problem that Schlenker's proposal (17) was faced with. To solve this problem, I'd like to submit that inferences triggered by DC do not receive the same epistemic treatment as root as presuppositions. It is a common assumption, following Stalnaker, that, at root, presuppositions are epistemically interpreted as in (25).

(25) Stalnaker's Bridge Principle. If a sentence ϕ presupposes that p, it can be felicitously used in context C only if C entails p.

[13] For a proof that the local context of the clause that is embedded under 'unaware' is the one given here, see [3]. $w' \in \text{DOX}^{w}_{\text{M.}}$ iff w' is compatible with what Mary believes in w.

I would like to propose that DC-triggered inferences are epistemically ambiguous in the following sense. Intuitively, for a sentence ϕ to be acceptable in context C, (25) requires that the presupposition of ϕ be true at every world of C. I would like to claim that DC-triggered inferences come with the following requirement: either every world of C makes the DC-triggered inferences true or every world of C *in which the assertive content of the sentence (without the gesture) is true* makes the DC-triggered inferences true. Let me implement this idea. Let W be the set of all possible worlds, and $\phi[\alpha_{\text{G}}]$ a sentence that contains a predicate-accompanying gesture. Construct the set C* such that (i) C* admits $\phi[\alpha_{\text{G}}]$ and (ii) no super-set of C* admits $\phi[\alpha_{\text{G}}]$. Then, $\phi[\alpha_{\text{G}}]$ can be felicitously used in a context C only if either $\text{C} \subseteq \text{C}^*$ or $(\text{C} \cap \{w : [\![\phi[\alpha]]\!]^w\}) \subseteq \text{C}^*$.

Let me briefly show why this move solves the problems of (24). Regarding the inference generated by (24a) in the scope of 'some but not all', we now have two options as to its epistemic treatment. Option one is that we impose the universal inference (that for each of the guys g, if g punished his son, he did so by slapping him) on the common ground, as we have been doing all along. This of course generates undesirable inferences regarding the guys who did not punish their son. Option two is to require the following: every world in the Context Set which makes the sentence 'some but not all of these ten guys punished his son' true, must make the inference that for each of the guys g, if g punished his son, he did so by slapping him true as well. This second option is a weaker imposition on the common ground than the first; for example, it is allowed that there be a world in the context set in which all guys punished their son by pulling his ear. What *is* required is that if some but not all guys punished their son, then all of them did so by slapping him, which is of course the target inference.

5 Conclusion

Co-speech gestures have only recently been studied by formal semanticists. Ebert & Ebert and Schlenker take a healthy attitude towards this freshly noticed phenomena: they try to assimilate them to better known phenomena (appositives in the case of E&E, presuppositions in the case of Schlenker) and study how they diverge. The attitude taken in this paper was to build on the disciplined approach of Schlenker in particular and ask the following question: what is the minimum amount of change that the cosuppositional analysis must go through, to make it empirically adequate? The resulting system is certainly rather baroque. My hope is that its empirical force can be used a basis to build a conceptually more elegant system.

References

[1] Cornelia Ebert and Christian Ebert. Gestures, demonstratives, and the attributive/referential distinction. Handout of a talk given at Semantics and Philosophy in Europe (SPE 7), Berlin, 2014.

[2] Philippe Schlenker. Local contexts. *Semantics and Pragmatics*, 2009.

[3] Philippe Schlenker. Gesture projection and cosuppositions. *Linguistics & Philosophy*, to appear.

Fatalism and the Logic of Unconditionals*

Justin Bledin

Johns Hopkins University, Baltimore, Maryland, USA
jbledin@jhu.edu

Abstract

In this paper, I consider a variant of the ancient Idle Argument involving so-called "unconditionals" with interrogative wh-antecedents. This new Idle Argument provides an ideal setting for probing the logic of these close relatives of *if*-conditionals, which has been comparatively underexplored. In the course of refuting the argument, I argue that contrary to received wisdom, many *wh*-conditionals are not properly speaking 'unconditional' in that they do not entail their main clauses, yet *modus ponens* remains valid for this class of expressions. I make these lessons formally precise in a semantic system that integrates recent decision-theoretic approaches to deliberative modals with ideas from inquisitive semantics. My larger aim is to challenge standard truth preservation views of logic and deductive argumentation.

1 The New Idle Argument

In this paper, I consider a new version of one of the oldest arguments in philosophy: the "Idle Argument" (also known as the "Lazy Argument").[1] This notorious argument survives in Cicero's *De Fato* (44BCE), where it is associated with the Stoic philosopher Chrysippus, and it also appears in Origen's *Contra Celsus* (248CE) (Bobzien 2001). In the modern era, the argument resurfaces in Dummett (1964) and is also discussed by Stalnaker (1975).

The new Idle Argument involves so-called "unconditionals" with interrogative wh-adjuncts; it is inspired by a structurally similar argument of Charlow (*ms.*) involving *if*-conditionals. The setting is London during WWII just as sirens sound warning of an approaching air raid. As you deliberate about whether to cut your supper short and go take shelter, the Fatalist (calm as ever) points out the following:

(1) If you are going to be killed in the raid, then you're better off staying where you are than taking precautions. (After all, if you *are* going to be killed, then you're going to be killed whether or not you take precautions.)

He then continues down the other fork:

(2) On the other hand, if you aren't going to be killed, then you're better off staying where you are than taking precautions. (After all, if you *aren't* going to be killed in the raid, then you aren't going to be killed even if you neglect to take precautions.)

Putting this together, the Fatalist infers this alternative unconditional:

(3) So, whether or not you are going to be killed, you're better off staying where you are than taking precautions.[2]

*Much thanks to Nate Charlow for conversations in Belgrade in Summer 2016 that led me to write this paper. Thanks also to Lucas Champollion, Ivano Ciardelli, Haoze Li, and Kyle Rawlins for helpful discussion.

[1] The "Idle/Lazy Argument" is best regarded as an umbrella term with a family of related arguments falling in its extension. I consider only one of these here.

[2] Note that the corresponding indicative conditional sounds terrible:

(i) ?? If you are going to be killed or not, you are better off staying where you are than taking precautions.

This violates what Ciardelli (2016b) calls "Zaefferer's rule": if the alternatives for the antecedent cover the context set, use the unconditional form; otherwise, the regular conditional form is required.

Detaching its consequent, he concludes:

(4) So look, you are better off staying where you are than taking precautions.

Not surprisingly, you sense something amiss with this argument, and so you set off towards the air-raid shelter. But why? What exactly is wrong with the new Idle/Lazy argument?

In §2-5, I consider several lines of response. I will be kind to the Fatalist—when ambiguity threatens, I will grant him readings necessary to make a premise hold or an inferential step go through. Running through the Idle Argument in this generous spirit, I want to see how far he can get. Spoiler alert: I ultimately conclude that the Fatalist can safely reach (3), but the final step of the argument is then problematic where he infers his conclusion (4) from this unconditional. I sharpen this diagnosis in §6, where I flesh out the context of the Idle Argument in more detail and present a formal semantics for *iffy better off* sentences that makes the essential features of my informal diagnosis formally precise. While the new Idle Argument reveals that many *wh*-conditionals do not entail their consequents, I conclude in §7 by arguing that *modus ponens* is nevertheless valid for these constructions.

2 Possible Escape Routes

Moving forward more carefully now, we can observe that the argument (1)-(4) relies on two prima facie plausible principles for unconditionals.

(5) **CA for *or not* unconditionals** **Consequent entailment (CE)**

$$\frac{\textit{If } \varphi, \psi \qquad \textit{If not-}\varphi, \psi}{\textit{Whether or not } \varphi, \psi} \qquad\qquad \frac{\textit{Whether or not } \varphi, \psi}{\psi}$$

Anyone looking to escape the Fatalist's conclusion must therefore respond in one of these ways:

(i) Reject one or both of the conditionals (1) and (2). I consider this option in §3.

(ii) Reject or restrict CA for *or not* unconditionals (this must be done only for readings of the indicative on which premises (1) and (2) both hold). More on this in §4.

(iii) Reject or restrict CE (for any reading of the unconditional (3) on which it follows from the Fatalist's premises). More on this in §5.

(iv) Play around with logic form. One might argue that we do not have a genuine instance of CA or CE on our hands.

(v) Take a desperate measure. For instance, one might deny the transitivity of entailment. I set options (iv) and (v) aside here.

3 On the Premises

In the half-century or so since the publication of Dummett's (1964) "Bringing About the Past", there has been an explosion of research on conditionals and modality. So there is now more room than ever to debate the Fatalist's premises. However, I'm willing to just grant the Fatalist his premises, for a couple of reasons.

First, there is at least one natural reading of the conditionals (1) and (2) on which they are difficult to deny. I submit that these premises have a "reflecting" reading (Cariani, Kaufmann & Kaufmann 2013) on which they are evaluated relative to the actual or potentially available information of some deliberating agent or agents (the natural choice: you) together with some representation of the agent's preferences and perhaps also a method for making decisions, and

by the time we get around to evaluating the embedded claim of comparative *betterness*, the background information state has been updated with the information that you will be killed (in the case of (1)) or won't be (in the case of (2)). On this reading, how can the conditionals be rejected? In the first instance where it is provisionally taken for granted that you will be killed, the choice between staying where you are and taking precautions is one between death and death after cutting your dinner short and trudging outside. In the second instance, the choice is between life and life with this bother. Either way, isn't it clearly better to stay put?

Admittedly, there are additional readings on which (1) and (2) do not sound nearly as good (Cariani et al.'s 2013 "non-reflecting" reading, for example). But rather than trying to argue that one or both of these premises fail to hold on *any* reading, let me also point out that there are structurally parallel arguments to the Idle Argument in §1 with fairly innocuous premises but terrible conclusions. Arguments like Missing Cat suggest that we do well to venture downstream from the premises of the Idle Argument and focus on its inferential steps:

Missing Cat. Grandma Rose has two orange tabbies and one gray shorthair. Grandma Pearl has two gray shorthairs and one orange tabby. Unfortunately, one of these cats has gone missing. Each of the cats is as likely to have gone missing as any of the others.

(6) If Grandma Rose lost one of her cats, then it is not equally likely that an orange or a gray cat went missing.

(7) Likewise, if Grandma Pearl lost one of her cats, then it is also not equally likely that an orange or a gray cat went missing.

(8) So, whether it was Grandma Rose that lost one of her cats or Grandma Pearl, it is not equally likely that an orange or a gray cat went missing.

(9) So, it is not equally likely that an orange or a gray cat went missing.

4 CA for Unconditionals

Going forward, I follow much of the literature on the semantics of questions in assuming that interrogatives can be assigned *alternative sets* (Hamblin 1973; Groenendijk & Stokhof 1984; Ciardelli et al. 2018). Crucially, I assume this holds for embedded clauses with interrogative morphology as well—in particular, I take it that the wh-adjuncts of alternative unconditionals contribute the same alternative sets as the corresponding root questions (Rawlins 2013; Ciardelli 2016b). For example, the antecedent of (8) introduces the two possible answers to the question it expresses: that Rose lost a cat, and that Pearl lost a cat.

Now, a common reaction to Missing Cat is to pin the blame on the inference from (6) and (7) to (8) using CA. These CA-rejectors seem to be evaluating the likelihood claim embedded in the unconditional (8) against a domain where the missing cat might be any of the six cats. This suggests the following interpretation strategy:

(10) **Flattened interpretation of alternative *wh*-conditionals**
An alternative unconditional *Whether* φ *or* ψ, χ is evaluated relative to an information state by first adjusting this state to support the information that at least one of the alternatives contributed by *Whether* φ *or* ψ holds and then evaluating χ with respect to the updated state.

In fact, alternative unconditionals are widely regarded to presuppose that one of the alternatives for their antecedent holds (more on this in §6). With felicitous uses, the initial update step in (10) is inert and we can evaluate *Whether* φ *or* ψ, χ simply by considering χ. The upshot: if unconditionals are interpreted along the lines of (10), then CA can fail but CE trivially holds.

However, the Fatalist might now argue that (10) isn't the right way to evaluate alternative unconditionals, or at least that (10) isn't the *only* way to evaluate them, and the "flattened" reading of (3) isn't what he had in mind regardless. Indeed, many semanticists working on unconditionals accept this alternative treatment (Rawlins 2013; a.m.o.):

(11) **Pointwise interpretation of alternative *wh*-conditionals**
Whether φ *or* ψ, χ is evaluated with respect to an information state by updating it with each of the alternatives for the antecedent in turn. If χ holds in each of the subordinate contexts induced by the different alternatives, then the unconditional holds.

So the idea is to evaluate (3) not by 'updating' with the tautology that you will be killed or not, but rather by first updating with the information that you will be killed and asking whether it is better to take precautions under this supposition, and then updating with the information that you will not be killed and asking about the precautions. If the value of (3) turns on whether both of these pointwise applications of the Ramsey Test pass, then (3) *does* seem to follow from (1) and (2). More generally, CA seems to fall directly out of (11) (as does its converse SDA).

The Fatalist can offer empirical considerations for thinking that alternative unconditionals are often (if not always) read pointwise. One kind of consideration is that *Whether* φ *or* ψ, χ is commonly used to send a stronger message than plain χ in a way predicted by (11) but not by (10), at least not straightforwardly. Compare the following:

(12) Whether Rodrigo or Brenda is making dinner, we might need to order takeout.
(13) We might need to order takeout.

Suppose the context is one in which it is taken for granted that Rodrigo or Brenda is making dinner (so the exhaustivity presupposition of (12) is met). In uttering (12), a speaker is arguably conveying that her current state of knowledge (or some other relevant body of information) leaves open both Rodrigo-makes-dinner possibilities and Brenda-makes-dinner possibilities in which disaster strikes and we need to order takeout. In contrast, one can utter (13) if Rodrigo or Brenda is an excellent cook, so long as the other is capable of ruining groceries.

The following examples further support the existence of pointwise readings:

(14) *Whether Julia is vacationing in Venezuela or Brazil, she might be in Caracas.
(15) *Whether or not Alfonso comes to the party, if Alfonso comes, you should come.

These sound not just false but absurd. However, this is surprising if the alternatives for the antecedents are flattened and both (14) and (15) are equivalent to their main clauses.

5 Consequent Entailment

I am suggesting that the Fatalist can get all the way to (3) by appealing to available readings for *if*-conditionals and *wh*-conditionals. Can he cross the final gap and reach his conclusion (4) using CE? No. This is, I think, where we should make our stand.

On what I have been calling the "reflecting" reading of (1), its value depends on what you're better off doing relative to information according to which you will die (together with a set of preferences, a decision rule, or whatever other structure is needed). The value of (2) likewise depends on information updated to support that you will survive. So, both premises presumably hold, as does (3) when interpreted pointwise, which rises or falls together with the conjunction of the Ramsey tests. However, the conclusion (4) presumably turns on what you're better off doing in your original non-updated information state, where you remain ignorant about whether you're going to killed and uncertain about what you're going to do, so this non-conditional claim doesnt hold. More generally, *betterness* claims are, like *likelihood* claims, highly sensitive to the information states against which they are evaluated. So CE can fail.

To be clear, I am *not* calling for a blanket rejection of the CE rule for *wh*-conditionals on their pointwise reading. Many (unflattened) *wh*-conditionals *do* entail their consequents:

(16) Whether it was Rose that lost one of her cats or Pearl, there's a fireman on the way.

(17) Whether Rodrigo or Brenda is making dinner, we're probably having pasta.

These sentences entail that there is a fireman coming and that we are probably having pasta for dinner. Unlike the consequents that have created trouble for CE, those of (16) and (17) are informationally 'well-behaved' (I sharpen the conditions under which CE is reliable in §6).

6 A Decision-Theoretic Semantics

Let us now assume that the logical forms of (1)-(4) can be represented at a suitable level of abstraction using a formal language $\mathcal{L}$ generated from a stock of atomic sentence letters $At_{\mathcal{L}}$, negation '$\neg$', conjunction '$\wedge$', and disjunction '$\vee$' in the usual way. The language $\mathcal{L}$ also includes a binary *better* operator '$\bigstar$' whose arguments are restricted to basic non-modal sentences built from the Boolean connectives, a question operator '?' whose single argument also takes only sentences in this basic fragment, and a conditional operator '$>$' whose first argument (antecedent) is restricted to basic sentences and basic sentences preceded by '?' but whose second argument (consequent) is unrestricted. Let $S_{\mathcal{L}}$ be the set of all sentences of $\mathcal{L}$.[3]

I interpret sentences in $S_{\mathcal{L}}$ with respect to *decision-theoretic* structures that encode (i) an agent's preferences over outcomes obtainable if she acts in certain ways and certain states of the world prevail, (ii) her information about these states, and (iii) her method of choosing between the options (see Carr 2012; Charlow 2016; Lassiter 2017 for related proposals). I call these structures "decision states". Their first component is a "decision problem":

(18) **Decision problems**
A *decision problem* DP over $\mathcal{W}$ is a tuple $\langle A, S, U, C\rangle$ where
a. $A, S \subseteq \mathcal{P}(\mathcal{W})$ are partitions of propositions (the *action set* and *state space*)
b. $U : A \times S \to \mathbb{R}$ maps action-state pairs to real numbers (the *utility function*)
c. $C : \mathcal{P}(\mathcal{W}) \to \mathbb{R}[0, 1]$ maps propositions to the unit interval (the *credence function*)
(I assume C is a probability measure over a finite space $\mathcal{W}$ in what follows.)

Their second component is a "decision rule" that evaluates the actions of decision problems:

(19) **Decision rules**
A *decision rule* R is a function that maps a decision problem DP to a partial order $\leq_{R(DP)}$ over its action set A.

If $a_1 \leq_{R(DP)} a_2$ then performing a_2 is at least as good as performing a_1 according to the rule R. For instance, rational agents might implement the following rule **MaxEU**:

(20) $a_1 \leq_{\textbf{MaxEU}(\mathrm{DP})} a_2$ iff $EU(a_1) \leq EU(a_2)$, where $EU(a) = \sum_{s \in S} C(s|a) \times U(a, s)$.[4]

But I don't want to insist that Expected Utility Theory has a monopoly on rational decision making, so I allow for other decision rules besides.

In the context of the Idle Argument, you face the dilemma of choosing between taking shelter or staying put. Suppose that the outcome of your decision depends on whether a bomb is dropped in your vicinity and, if so, its size. If a large bomb is dropped, you're dead either way. If no bomb is dropped, you survive either way. But if a small bomb is dropped, then you live iff you take cover. This DP—call it **Air Raid**—has the following action set/state space:

[3] 'K', 'S', and 'P' abbreviate 'You are going to be killed', 'You stay where you are', and 'You take precautions' respectively, φ_0, ψ_0, ... range over sentences in the basic fragment of $\mathcal{L}$, and φ, ψ, ... range over all sentences.

[4] I work with this version of Expected Utility Theory for ease of exposition. I'm not looking to take a stand between causal vs. evidential decision theory.

(21) $A_{\mathbf{AR}} = \{\lambda w_s.$you take precautions in $w, \lambda w_s.$you stay where you are in $w\}$
(22) $S_{\mathbf{AR}} = \{\lambda w_s.$large bomb in $w, \lambda w_s.$small bomb in $w, \lambda w_s.$no bomb in $w\}$

To further fix ideas, suppose that the value of extended life is 100 utils, death is valued at -100 utils, and making efforts contributes a relatively minor loss of a single util:

(23) $U_{\mathbf{AR}}(\lambda w.$you take precautions in $w, \lambda w.$large bomb in $w) = -101$
$U_{\mathbf{AR}}(\lambda w.$you take precautions in $w, \lambda w.$small bomb in $w) = 99$ etcetera.

Furthermore, suppose you have these conditional credences:

(24) $C_{\mathbf{AR}}(s|a) = 1/3$ for each $a \in A$ and $s \in S$.

In this case, a simple calculation establishes that the expected utility of taking precautions is greater than that of staying where you are:

(25) $\mathcal{V}(S) <_{\textbf{MaxEU(Air Raid)}} \mathcal{V}(P)$

So Expected Utility Theory recommends ignoring the Fatalist.

We can now simultaneously assign support $\models$ and reject $=\!|$ conditions to the sentences in $S_{\mathcal{L}}$ parametrized to the following structures:

(26) **Decision states**
A *decision state* $d = \langle DP_d, R_d\rangle$ consists of a decision problem and decision rule.

Sentence letters are supported by decision states whose DP-parameter includes a credence function all of whose mass is concentrated on the α-worlds in $\mathcal{W}$ and rejected by states whose credal mass is spread entirely across not-α-worlds:

(27) **Interpretation of atomic formulae**
$d \models \alpha$ iff $C_{DP_d}(\mathcal{V}(\alpha)) = 1$ $\quad$ $d =\!| \alpha$ iff $C_{DP_d}(\mathcal{V}(\alpha)) = 0$

Negation flips between support and rejection (Hawke & Steinert-Threlkeld 2016):

(28) **Interpretation of negation**
$d \models \neg\varphi$ iff $d =\!| \varphi$ $\quad$ $d =\!| \neg\varphi$ iff $d \models \varphi$

Conjunction and disjunction are defined as follows (cf. Ciardelli et al. 2018):

(29) **Interpretation of conjunction and disjunction**
$d \models \varphi \wedge \psi$ iff $d \models \varphi$ and $d \models \psi$ $\quad$ $d =\!| \varphi \wedge \psi$ iff $d =\!| \varphi$ or $d =\!| \psi$
$d \models \varphi \vee \psi$ iff $d \models \varphi$ or $d \models \psi$ $\quad$ $d =\!| \varphi \vee \psi$ iff $d =\!| \varphi$ and $d =\!| \psi$

To interpret *better*, we must first introduce another notion of *propositional support* for sentences in $S_{\mathcal{L}}$ relative to qualitative information states, modeled as sets of possible worlds. Every decision state d determines such an information state consisting of the worlds in $\mathcal{W}$ assigned nonzero probability by its component credence function:

(30) $i_d = \{w \in \mathcal{W} : C_{DP_d}(\{w\}) > 0\}$.

Propositional support is defined in terms of these states:

(31) **Propositional support**
Given any information state $i \subseteq \mathcal{W}$ and $\varphi \in S_{\mathcal{L}}$, $i \models \varphi$ iff for any d s.t. $i = i_d$, $d \models \varphi$.

Note for example that

(32)
$i \models P \vee S$ iff for any d such that $i = i_d$, $d \models P \vee S$
iff for any..., $C_{DP}(\mathcal{V}(P)) = 1$ or $C_{DP}(\mathcal{V}(S)) = 1$
iff $i \subseteq \mathcal{V}(P)$ or $i \subseteq \mathcal{V}(S)$.[5]

[5]This is the support condition from the most basic system of inquisitive semantics, `InqB` (Ciardelli et al. 2018). The propositional support conditions in (31) coincide with those in `InqB` for basic sentences without negation, but they can diverge for negated sentences. For instance, $i \models \varphi_0$ iff $i \models \neg\neg\varphi_0$ in our system but this equivalence fails in `InqB`. The differences between the systems don't matter for present purposes.

As in inquisitive semantics, we next define the *alternatives* for $\varphi \in S_{\mathcal{L}}$ to be the maximal (qualitative) information states that support it (Ciardelli et al. 2018):

(33) $alt(\varphi) = \{i \subseteq \mathcal{W} : i \models \varphi \text{ and there is no } i \subset i' \text{ s.t. } i' \models \varphi\}$

For example, the set of alternatives for $P \vee S$ is $alt(P \vee S) = \{\mathcal{V}(P), \mathcal{V}(S)\}$.

The semantics for *better* is defined in terms of these alternative sets:

(34) **Interpretation of comparative betterness**
$d \models \bigstar(\varphi_0, \psi_0)$, $d \mathrel{=\!\!|} \bigstar(\varphi_0, \psi_0)$ are defined only if $alt(\varphi_0), alt(\psi_0) \subseteq A_{DP_d}$.[6] If defined,

$d \models \bigstar(\varphi_0, \psi_0)$ iff for all $a \in alt(\varphi_0)$ and $a' \in alt(\psi_0)$, $a' <_{R_d(DP_d)} a$.
$d \mathrel{=\!\!|} \bigstar(\varphi_0, \psi_0)$ iff for some $a \in alt(\varphi_0)$ and $a' \in alt(\psi_0)$, $a \leq_{R_d(DP_d)} a'$.

To get a feel for (34), consider the following argument (Lassiter's 2017 "Disjunctive Inference"):

(35) It is better to mail the letter than to burn it.
(36) It is better to mail the letter than to throw it in the trash.
(37) So, it is better to mail the letter than to either burn it or throw it in the trash.

Our semantics nicely predicts that this reasoning is impeccable. Translating the argument as $\bigstar(M, B), \bigstar(M, T) \therefore \bigstar(M, B \vee T)$, (34) implies that d supports (35) iff the set of M-worlds in which you mail the letter and the set of B-worlds in which you burn it are both actions of DP_d and R_d recommends the former over the latter. Similarly, d supports (36) iff the set of T-worlds in which you throw the letter away is also an action of DP_d and R_d recommends mailing the letter over discarding it. But then d must also support (37), as the alternatives for M and the "inquisitive" $B \vee T$ are all actions and the former is preferred to each of the latter actions.[7]

To formalize the Fatalist's premises, we still need a semantics for '$>$'. The intuitive idea behind my proposal is that a decision state d supports a conditional of the form $\varphi_0 > \psi$ or $?\varphi_0 > \psi$ iff every way of minimally updating DP_d with one of the alternatives for φ_0 or $?\varphi_0$ delivers a decision state that supports the consequent ψ (this semantics is inspired by related proposals in Yalcin 2007; Kolodny & MacFarlane 2010; Ciardelli 2016b; a.o.). This set of updated states $d \oplus \varphi_{?0}$ is determined as follows, where $\varphi_{?0}$ is a sentence of the form φ_0 or $?\varphi_0$:

(38) $DP + i = \langle A_{DP}, S_{DP}, U_{DP}, C_{DP}(\cdot|i)\rangle$
Defined only if $C_{DP}(i) > 0$.

(39) $d \oplus \varphi_{?0} = \{d' : d' = \langle DP_d + i, R_d\rangle \text{ for some } i \in alt(\varphi_{?0})\}$

With (39) in hand, conditional expressions can now be evaluated in this Ramseyian manner:

(40) **Interpretation of conditional operator**
$d \models \varphi_{?0} > \psi$, $d \mathrel{=\!\!|} \varphi_{?0} > \psi$ are defined only if $d \models \varphi_{?0}$ is defined and for all $d' \in d \oplus \varphi_{?0}$, $d' \models \psi$ is defined. If defined,

$d \models \varphi_{?0} > \psi$ iff for all $d' \in d \oplus \varphi_{?0}$, $d' \models \psi$
$d \mathrel{=\!\!|} \varphi_{?0} > \psi$ iff for some $d' \in d \oplus \varphi_{?0}$, $d' \not\models \psi$

The definability condition in (40) ensures that presuppositions project out of the antecedents of conditionals. This is important when proving various facts about the logic of *wh*-conditionals.

Applying (40) to premise (1) of the Idle Argument gives us:

(41) $d \models K > \bigstar(S, P)$ iff for all $d' \in d \oplus K$, $d' \models \bigstar(S, P)$
iff $\mathcal{V}(P) <_{R_d(DP_d + \mathcal{V}(K))} \mathcal{V}(S)$ (assuming presuppositions met)

[6] We take it to be a presupposition of an action-guiding sentence of the form $\bigstar(\varphi_0, \psi_0)$ that the alternatives for each argument are actions of the DP against which it is evaluated. When I say that "$d \models \varphi$ is (un)defined" or "$d \mathrel{=\!\!|} \varphi$ is (un)defined", what I really mean to say is that the characteristic function for the relation $\models$ or $\mathrel{=\!\!|}$ is (un)defined on $\langle d, \varphi\rangle$. With presuppositions around, these characteristic functions are partial.

[7] I assume here that valid (deductively good) inferences preserve support. More on this in a few paragraphs.

So if we assume that

(42) $C_{\mathbf{AR}+\mathcal{V}(K)}(\lambda w_s.\text{large bomb in } w|\lambda w_s.\text{precautions in } w) = 1$
$C_{\mathbf{AR}+\mathcal{V}(K)}(\lambda w_s.\text{large bomb in } w|\lambda w_s.\text{stay put in } w) = 1/2$
$C_{\mathbf{AR}+\mathcal{V}(K)}(\lambda w_s.\text{small bomb in } w|\lambda w_s.\text{stay put in } w) = 1/2$

this premise is supported by $\langle$**Air Raid**, **MaxEU**$\rangle$. Making similar assumptions, one can also establish that premise (2) is supported by this state. The semantics thus allows us to see how both of the Fatalist's premises can hold with respect to a single decision state—at least when these premises are understood "reflectively".

How does the rest of the Idle Argument play out? To evaluate the unconditional (3), we need to round the semantics off with an entry for '?'. I follow Rawlins (2013) in assuming that the sole function of the question operator is to contribute new presuppositions to the effect that one and only one alternative for the basic sentence it operates on holds:

(43) **Interpretation of question operator**
$d \models ?\varphi_0, d =| ?\varphi_0$ are defined only if (i) $C_{DP_d}(\bigcup alt(\varphi_0)) = 1$, and (ii) for all $i, i' \in alt(\varphi_0)$ where $i \neq i'$, there is no $w \in i \cap i'$ such that $C_{DP_d}(\{w\}) > 0$.
If defined, $d \models ?\varphi_0$ iff $d \models \varphi_0$ and $d =| ?\varphi_0$ iff $d =| \varphi_0$.[8]

The exhaustivity (i) and exclusivity (ii) constraints in (43) project out of the *wh*-adjunct of alternative unconditionals, which thereby presuppose that exactly one alternative for their antecedent holds. For the special case of *or not wh*-conditionals, these presuppositions are trivially satisfied. When evaluating (3), the question operator can be ignored:

(44) $d \models ?(K \vee \neg K) > \bigstar(S, P)$
iff for all $d' \in d \oplus (K \vee \neg K)$, $d' \models \bigstar(S, P)$
iff $\langle DP_d + \mathcal{V}(K), R_d\rangle \models \bigstar(S, P)$ and $\langle DP_d + \mathcal{W}\backslash\mathcal{V}(K), R_d\rangle \models \bigstar(S, P)$
iff $d \models K > \bigstar(S, P)$ and $d \models \neg K > \bigstar(S, P)$.

So relative to the **MaxEU** rule at least, the Fatalist's use of CA doesn't lead him astray. However, given the earlier result (25), his conclusion (4) is rejected by $\langle$**Air Raid**, **MaxEU**$\rangle$.

To assess the validity of the inference rules CA or CE themselves, we still need to define a formal notion of consequence over $\mathcal{L}$. Because we are working with both support and reject conditions, there are a number of different options. Leaving a more detailed exploration of these options for the future, we simply require that whenever support conditions are defined for the premises and conclusion of an argument, this argument preserves support (this is basically what you get by crossing a decision-theoretic upgrade of Yalcin's 2007 "informational consequence" (see also Veltman's 1996 '$\models_3$') with von Fintel's 1999 "Strawson-entailment"):

(45) **Strawsonian support-preserving consequence**
$\{\varphi_1, ..., \varphi_n\} \models \psi$ iff for any decision state d such that $d \models \varphi_1, ..., d \models \varphi_n, d \models \psi$ are defined, if $d \models \varphi_1, ..., d \models \varphi_n$, then $d \models \psi$.

It can be shown that the general CA rule for alternative *wh*-conditionals is validated by (45):

(46) **CA is valid.** $\{\varphi_0 > \chi, \psi_0 > \chi\} \models ?(\varphi_0 \vee \psi_0) > \chi$.[9]

However, CE isn't unrestrictedly valid, as discussed above:

[8] This is another place where I diverge from inquisitive semanticists, who treat '?' as a kind of projection operator definable in terms of negation and disjunction: $?\varphi_0 := \varphi_0 \vee \neg\varphi_0$ (Ciardelli et al. 2018).

[9] It is crucial that support for the conclusion is defined; if not, CA needn't preserve support. To see this, consider a state d such that $i_d = \{w_1, w_2, w_3\}$, where w_1 is the only A-world, w_2 is the only B-world, w_3 is the only C-world, and all three worlds are D-worlds. Although $d \models A > D$ and $d \models B > D$, $d \models ?(A \vee B) > D$ is undefined because the exhaustivity presupposition contributed by its '?'-adjunct isn't satisfied.

(47) **CE is invalid.** $\{?(\varphi_0 \vee \psi_0) > \chi\} \not\models \chi$.

As mentioned in §4, alternative *wh*-conditionals still entail their main clauses in a broad range of cases. Let us call a sentence $\varphi \in S_{\mathcal{L}}$ "coarsely distributive" iff it has the following property:

(48) **Coarse distributivity**
The sentence $\varphi \in S_{\mathcal{L}}$ is *coarsely distributive* iff for any partition $I = \{i_1, ..., i_n\}$ over $\mathcal{W}$ and state d, if $\langle DP_d + i_1, R_d\rangle \models \varphi$,..., and $\langle DP_d + i_n, R_d\rangle \models \varphi$, then $i \models \varphi$.[10]

CE is valid so long as we restrict our attention to such sentences:

(49) **CE is valid for coarsely distributive consequents.**
For any coarsely distributive χ, $\{?(\varphi_0 \vee \psi_0) > \chi\} \models \chi$.

As also mentioned in §4, CE holds for alternative *wh*-unconditionals if these receive a flattened interpretation. We can recover this reading in our system by adding a 'flattening' operator '!' to $\mathcal{L}$ that can be inserted before basic sentences and basic sentences preceded by '?':[11]

(50) **Flattening operator**
$d \models !\varphi_{?0}, d \neq !\varphi_{?0}$ are defined only if $d \models \varphi_{?0}$ is defined. If defined,

$$d \models !\varphi_{?0} \quad \text{iff} \quad C_{DP_d}(\bigcup alt(\varphi_{?0})) = 1$$
$$d \neq !\varphi_{?0} \quad \text{iff} \quad C_{DP_d}(\bigcup alt(\varphi_{?0})) < 1$$

If (40) is extended in the natural way to accommodate conditionals with $!\varphi_{?0}$-antecedents, CE holds for the flattened case:

(51) **CE is valid for flattened *wh*-conditionals.**
$\{!?(\varphi_0 \vee \psi_0) > \chi\} \models \chi$.

7 Is *Modus Ponens* Valid for *Wh*-conditionals?

To conclude, I want to say a few words about how *not* to conclude. Charlow (*ms.*) argues on the basis of similar arguments to the Idle Argument in §1 that *modus ponens* (MP) is invalid. But while the adjunct of (3) might seem tautological, it is incorrect to think we have a failure of MP here. Crucially, the antecedent of (3) is the *interrogative* sentence 'Whether or not you are going to be killed'. More generally, MP for *or not wh*-conditionals takes the following form:

(52) **MP for *or not* unconditionals**

$$\frac{\textit{Whether or not } \varphi, \chi \qquad \textit{Whether or not } \varphi}{\psi}$$

In fact, if we grant the Fatalist the extra premise $?(K \vee \neg K)$, then his argument goes through. It can be shown that the following general MP rule for alternative *wh*-conditionals is valid:

(53) **MP is valid.** $\{?(\varphi_0 \vee \psi_0) > \chi, ?(\varphi_0 \vee \psi_0)\} \models \chi$.

So in particular $?(K \vee \neg K) > \bigstar(S, P)$, $?(K \vee \neg K) \therefore \bigstar(S, P)$. Note, however, that all the extended argument establishes is that a decision state that supports (3) and *settles* the question of whether or not you will be killed also supports that it is better to stay where you are than to take precautions (assuming that support is even defined; see Ciardelli 2016a for further helpful discussion about argumentation with questions). In other words, the extended fatalistic argument establishes only that you are better off staying put when it is known what will come to pass—hardly a result that will lead the youth to a life of idleness.

[10] Are all basic non-modal sentences coarsely distributive? No. Simple disjunctions like $A \vee B$ fail to distribute.

[11] A similar flattening operator appears in inquisitive semantics but it is there defined in terms of double negation: $!\varphi_0 := \neg\neg\varphi_0$ (Ciardelli et al. 2018). This clearly won't work in our system because—as mentioned in n. 5 above—negations cancel each other out.

References

Susanne Bobzien. *Determinism and Freedom in Stoic Philosophy*. Oxford University Press, Oxford, 2001.

Fabrizio Cariani, Magdalena Kaufmann, and Stefan Kaufmann. Deliberative modality under epistemic uncertainty. *Linguistics and Philosophy*, 36(3):225–259, 2013.

Jennifer Carr. Deontic modals without decision theory. In Emmanuel Chemla, Vincent Homer, and Grégoire Winterstein, editors, *Proceedings of Sinn und Bedeutung 17*, pages 167–182, 2012.

Nate Charlow. Decision theory: Yes! Truth conditions: No! In Nate Charlow and Matthew Chrisman, editors, *Deontic Modality*, pages 47–81. Oxford University Press, 2016.

Nate Charlow. Another counterexample to Modus Ponens. Unpublished manuscript.

Ivano Ciardelli. *Questions in Logic*. Ph.D. dissertation, University of Amsterdam, 2016a.

Ivano Ciardelli. Lifting conditionals to inquisitive semantics. In Mary Moroney, Carol-Rose Little, Jacob Collard, and Dan Burgdorf, editors, *Proceedings of SALT 26*, pages 732–752, 2016b.

Ivano Ciardelli, Jeroen Groenendijk, and Floris Roelofsen. *Inquisitive Semantics*. Oxford University Press, Oxford, 2018.

Michael Dummett. Bringing about the past. *Philosophical Review*, 73(3):338–359, 1964.

Jeroen Groenendijk and Martin Stokhof. *Studies on the semantics of questions and the pragmatics of answers*. Ph.D. dissertation, University of Amsterdam, 1984.

C. L. Hamblin. Questions in Montague English. *Foundations of Language*, 10(1):41–53, 1973.

Peter Hawke and Shane Steinert-Threlkeld. Informational dynamics of epistemic possibility modals. *Synthese*, 2016.

Niko Kolodny and John MacFarlane. Ifs and oughts. *Journal of Philosophy*, CVII(3):115–143, 2010.

Daniel Lassiter. *Graded Modality: Qualitative and Quantitative Perspectives*. Oxford University Press, Oxford, 2017.

Kyle Rawlins. (Un)conditionals. *Natural Language Semantics*, 21(2):111–178, 2013.

Robert Stalnaker. Indicative conditionals. *Philosophia*, 5(3):269–286, 1975.

Frank Veltman. Defaults in update semantics. *Journal of Philosophical Logic*, 25(3):221–261, 1996.

Kai von Fintel. NPI licensing, Strawson entailment, and context dependency. *Journal of Semantics*, 16(2):97–148, 1999.

Seth Yalcin. Epistemic modals. *Mind*, 116(464):983–1026, 2007.

Splitting Germanic negative indefinites *

Dominique Blok[1], Lisa Bylinina[2], and Rick Nouwen[1]

[1] Utrecht University,
Utrecht, the Netherlands
[2] Leiden University,
Leiden, the Netherlands

Abstract

Constructions with an intensional verb and the negative indefinite *geen* in Dutch (as well as *kein* in German) routinely lead to split scope readings. English *no* does not systematically give rise to such readings. Observing a number of other differences between *geen* / *kein* and *no*, we claim that there are two kinds of negative indefinites in Germanic: (i) degree quantifiers that consist of a negative and a numeral meaning component and give rise to split scope (Dutch *geen*, German *kein*); (ii) non-degree negative indefinites (English *no*, and its counterparts in e.g. Swedish). We argue that the split scope phenomenon is tied to degree quantifier movement and is essentially a degree phenomenon.

1 Split scope

Negative indefinites in Dutch and German are known to give rise to so-called **split scope readings** – the meaning of the negative indefinite seems to be split in two pieces by another scope-bearing element (Jacobs, 1980; Kratzer, 1995; Geurts, 1996; de Swart, 2000; Penka and Zeijlstra, 2005; Abels and Martí, 2010; Penka, 2011), illustrated here with universal and existential modals in Dutch:

(1) Je hoeft **geen** stropdas te dragen.
you must-NPI GEEN tie to wear
'You do not have to wear a tie.' $\neg > \Box > \exists$

(2) Henk mag **geen** toetje eten.
you may GEEN dessert eat
'Henk is not allowed to eat a dessert.' $\neg > \Diamond > \exists$

In this paper we are concerned with the nature of split scope. The standard quantifier semantics for negative indefinite determiners (including *no*, *geen* etc.), as in (3), does not straightforwardly split and, as such, it does not offer a straightforward account of the splitting phenomenon.

(3) $[\![\text{geen}]\!] = [\![\text{no}]\!] = \lambda P_{\langle et\rangle}\lambda Q_{\langle et\rangle}.P \cap Q = \varnothing$

As we will argue, whatever analysis substitutes (3) in order to allow for split scope, it should cover the following four observations we will make in this paper: (1) Split scope with negative indefinites is not generally available cross-linguistically; (2) Split scope with degree expressions *is* generally available cross-linguistically; (3) Split scope is constrained by a scope constraint

*Blok and Nouwen gratefully acknowledge a grant from the European Research Council under the European Unions Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement no. 313502. Bylinina gratefully acknowledges a grant from the Netherlands Organisation for Scientific Research / VENI Grant no. 275-70-045. Thanks are due to Eddy Ruys for helpful comments at an earlier stage of this research.

observed for degree expressions; (4) Negative indefinites that can modify numerals systematically allow for split scope readings. We will offer an analysis that rests on these observations, arguing that: (i) split scope is a degree phenomenon; (ii) Dutch *geen* and German *kein* are degree quantifiers, while English *no* isn't.

Note that on our proposal, Dutch *geen* and German *kein* are not indefinite determiners, but rather degree quantifiers. In what follows, we will nevertheless keep the descriptive label *indefinite* for these expressions. The reader should bear in mind that this label carries no theoretical commitment.

2 Properties of split scope

2.1 Split scope: Cross-linguistic limitations

Most studies of split scope with negative indefinites concern Dutch or German. Yet, split scope is sometimes discussed for English *no* (Potts, 2000; von Fintel and Iatridou, 2007; Iatridou and Sichel, 2011; Kennedy and Alrenga, 2014), usually illustrated with examples as the following:

(4) The company need fire **no** employees.
'It is not the case that the co. is obligated to fire an employee.' $\neg > \Box > \exists$

However, the phenomenon is much more restricted in English than in Dutch/German. Changing an NPI *need* to a neutral *have to* leads to the loss of the split scope reading:

(5) The company has to fire **no** employees.
'#It's not the case that the company has to fire an employee.' $\neg > \Box > \exists$

Similarly, a direct translation of the paradigmatic split scope example (1) into English results in a sentence with no split scope reading. It only has a *de dicto* reading.

(6) At this party, you have to wear no tie.

We take this to mean that English *no* lacks the *general* scope splitting ability of Dutch *geen*. This discrepancy will play a large role in our story below.

2.2 Split scope beyond negative indefinites

Apart from negative indefinites, *degree expressions* tend to split their scope (e.g. Hackl 2000). Importantly, they do so to the same extent in English as in Dutch / German:

(7) Tom has to bring **at most two** blankets.
'Tom does not have to bring more than two blankets' $\neg > \Box > > 2$

(8) They are allowed to write **few** letters.
'It is not the case that they are allowed to write many letters' $\neg > \Diamond >$ many

It is important to note several things here. First, all quantifiers in these examples are degree quantifiers. At first sight, degree quantifiers do not seem to form a natural class with *geen*-type expressions (or with *no*, for that matter). Why this particular collection of expressions (degree quantifiers + *geen* / *kein*) gives rise to split scope is a puzzle that our analysis will eliminate by giving *geen* / *kein* a semantics of a degree quantifier. Finally, in contrast to the behaviour of *no* that we observed in the previous subsection, split scope with English degree quantifiers is unlimited. That is, for both English and Dutch/German, degree quantifiers always have the

ability to split scope. The ability of negative indefinites to split scope is general for Dutch and German and severely limited for the case of English. The analysis we will develop below deals with this variation in a straightforward way: by treating split scope as a degree phenomenon and analyzing *geen* / *kein* as degree quantifiers, unlike *no*.

This kind of analysis has immediate appeal due to the fact that split scope readings with degree quantifiers come naturally under a relatively standard analysis of degree quantification, which we adopt here. According to this analysis, quantifiers like *at most n*, *fewer than n* and *few* are not type $\langle\langle e,t\rangle,\langle\langle e,t\rangle,t\rangle\rangle$ quantifiers, rather they are type $\langle\langle d,t\rangle,t\rangle$ (Hackl, 2000; Nouwen, 2008, 2010; Kennedy, 2015) with the kind of meaning shown in (9) for *at most two*. (Also note that under this analysis a silent MANY is needed to mediate the relation between the degree and the noun, see Hackl 2000 and below for more details). Given this analysis, split scope readings with degree quantifiers are straightforward cases of QR:

(9) $[\![\text{at most } 2]\!] = \lambda P_{\langle dt\rangle}.max(P) \leq 2$

(10) $[\![$ at most 2 [Tom has to bring ~~at most 2~~ MANY books] $]\!]$
$= [\![\text{at most } 2]\!](\lambda n.\, \Box\, \exists x\, [\textbf{*bring}(\textbf{T},x)\ \&\ \textbf{*book}(x)\&\#x = n])$
$= max(\{n|\, \Box\, \exists x[\textbf{*bring}(\textbf{T},x)\ \&\ \textbf{*book}(x)\&\#x = n]\}) \leq 2$

(11) $[\![\text{few}]\!] = \lambda P_{\langle dt\rangle}.max(P) < d_{st}$

If, as is standardly assumed, *geen*-type negative indefinites are not degree quantifiers, then an analysis of the split scope readings they give rise to will have to be quite different from what is illustrated in (10). That is, split scope will have to be essentially different in nature for degree quantifiers on the one hand and *geen* / *kein* on the other. Naturally, that would make it harder to explain their similar properties.

2.3 Split scope and the Heim-Kennedy generalization

We have seen modal verbs (*must*, *need*, *can*, *may*) split scope of *geen*-type indefinites. Are modals the only scope-splitters? With normal intonation, *geen*-type indefinites do not split scope over non-modal quantifiers. The following example from German illustrates this:

(12) Genau ein Arzt hat **kein** Auto.
exactly one doctor has KEIN car
#'It's not the case that exactly one doctor has a car'
'Exactly one doctor has no car'

The distribution of split scope is reminiscent of the Heim-Kennedy generalization (Kennedy, 1997; Heim, 2000): degree quantifiers can scope above (at least some) intensional verbs (14), but nominal quantifiers can never intervene between a degree quantifier and its trace (15).[1]

(13) *$[\text{D}_{dtt}\ \ldots\text{Q}_{ett}\ \ldots t_d]$

(14) Tom needs at most two blankets.
'Tom does not need more than three blankets.'

(15) Every student has at most three books.
'#Not every student has more than three books.'

Negative indefinites behave in a parallel fashion (example from Dutch):

[1] See Nouwen and Dotlačil (2017) for discussion of details as to how this constraint should be stated.

(16) Iedere student heeft **geen** oplossing gevonden.
every student has GEEN solution found
#'Not every student found a solution'

Why would split scope with *geen* obey a generalisation concerning degree quantifiers if it's not a degree quantifier? Once more, the data suggests that the broad phenomenon of scope splitting, including the splitting of negative indefinites, is a degree phenomenon.

2.4 *Geen*-type negative indefinites with numerals

We have seen above (Section 2.1) that there is a difference between *geen* / *kein* and *no* in that split scope is systematic with the former and restricted with the latter:

(1) Je hoeft **geen** stropdas te dragen.
you must-NPI GEEN tie to wear
'You do not have to wear a tie.' $\neg > \Box > \exists$

(17) At this party, you have to wear **no** tie. $*\neg > \Box > \exists$

We observe another difference between *geen* / *kein* and *no* – namely that *geen* / *kein* combine with numerals while *no* generally doesn't:

(18) Nigella heeft **geen** 20 taarten gebakken.
Nigella has GEEN 20 cakes baked.
'Nigella has not baked 20 cakes.'

(19) *Nigella baked no 20 cakes.

We suggest that this difference is not accidental, both cross-linguistically and semantically. A quick exploration of Germanic languages supports the following generalisation, which we call **the numeral modifier generalisation for negative indefinites in Germanic:** *whenever a negative indefinite can modify numerals, its capacity to create split scope readings with intensional operators is unlimited.*

We found that Icelandic and Frisian pair with Dutch and German in that they have negative indefinites (*eng* and *gjin*, respectively) which can modify numerals and which have unlimited split scope. The Swedish negative indefinite *ing* is like English: it lacks a use as a numeral modifier and does not generally give rise to split scope readings.

These differences, we believe, can help us point in the direction of an analysis of split scope readings of *geen*-type indefinites and the lack of such readings with *no*. In short, we suggest that *geen* is a degree quantifier, quite like other expressions subject to split scope. We first spell out an analysis of 'geen' in combination with numerals, as in (18), and then move on to the paradigmatic bare cases.

3 Analysis

3.1 *Geen* with numerals

Let's first implement the idea of *geen* as a degree quantifier by analysing cases like (18), where *geen* combines with a numeral. Sentences like (18) are ambiguous between a lower and a doubly bounded reading. Correspondingly, we propose that *geen* in construction with numerals comes in two guises, both expressing a particular form of scalar negation:

(20) $[\![\text{geen}_{=}]\!] = \lambda n_d \lambda P_{\langle dt\rangle}.\neg max(P) = n$

(21) $[\![\text{geen}_{\geq}]\!] = \lambda n_d \lambda P_{\langle dt\rangle}.\neg P(n)$

Both these senses of *geen* combine with a numeral of type d (degree) and a degree predicate – but with a somewhat different result.

(22) $[\![\text{N. baked } \mathbf{geen}_{=} \text{ 20 cakes}]\!] = \neg max\{n|\exists x[\mathbf{*baked}(\mathbf{N}, x) \;\&\; \mathbf{*cake}(x) \;\&\; \#x{=}n]\} = 20$

(23) $[\![\text{N. baked } \mathbf{geen}_{\geq} \text{ 20 cakes}]\!] = \neg\exists x[\mathbf{*baked}(\mathbf{N}, x) \;\&\; \mathbf{*cake}(x) \;\&\; \#x{=}20]$

(22) is true when the quantity of cakes that Nigella made is not twenty (it could be five or fifty or, in fact, zero – see below). (23) is true when Nigella baked fewer than twenty cakes. These are exactly the interpretations that are attested for (18).

These readings arise by following standard assumptions for the semantics of numerals and degree quantification. First, we assume that the numeral has semantic type d and forms a constituent with *geen* ('geen$_{=}$' and 'geen$_{\geq}$') in much the same way as a numeral modifier like *at least* combines with a numeral. We also assume a silent MANY, as in Hackl (2000) and much of the subsequent literature, which occupies the position between the numeral and the noun:[2]

(24) $[\![\text{MANY}]\!] = \lambda n_d \lambda P_{\langle e,t\rangle} \lambda Q_{\langle e,t\rangle}.\exists x[\#x = n \;\&\; \mathbf{*}P(x) \;\&\; \mathbf{*}Q(x)]$

Geen 20 QRs in order to resolve a type clash (as it is of type $\langle dt, t\rangle$ rather than d), leaving behind a trace of type d and creating the following degree predicate, with which *geen 20* will combine:

(25) $[\![\text{Nigella baked } n \text{ MANY cakes}]\!] = \lambda n_d.\exists x[\mathbf{*baked}(\mathbf{N}, x) \& \mathbf{*cake}(x) \& \#x = n]$

This set contains numbers such that it's true that Nigella baked at least this number of cakes.

After *geen 20* combines with (25), the meaning will depend on whether it is an 'exactly' ('=') version of *geen* or the 'at least' ('≥') version. The 'exactly'-version of *geen* ('geen$_{=}$') will then state that the maximal element of this set of degrees is not 20. 'At least' *geen* ('geen$_{\geq}$') will state that this set does not contain 20.

What about zero cakes ($\#x = 0$)? Sentences like (18) are true in a situation when Nigella baked nothing. To make sure our analysis predicts that, we spell out our assumptions about the structure of the plural domain. Following (Landman, 2011; Bylinina and Nouwen, 2017) a.o., we assume the bottom element $\perp$ is in the denotation of pluralised predicates $\mathbf{*P}$. That is, the domain of entities contains atoms and pluralities, including the zero plurality, the entity with cardinality 0. In other words, the domains are as illustrated in figure 1, where the atoms are in bold.

This semantics for plurals ensures that both 'geen$_{=}$ 20' and 'geen$_{\geq}$ 20' are compatible with the $\#x = 0$ alternative being true. (It also ensures that other downward entailing modified numerals are compatible with $\#x = 0$, which provides extra motivation for this particular setup. See Buccola and Spector (2016), Bylinina and Nouwen (2017) for discussion.)

Let's now turn to split-scope environments, where *geen* is embedded under a modal. In such an environment, the split scope reading is derived by *geen 20* QR-ing over the modal verb in a straightforward way:

[2] Note that the $\langle e, t\rangle$ arguments of MANY are pluralised. The syntactic details of this are beyond the immediate scope of this paper, but we believe the differences between DPs like *one book* and *two books* do not reside in the semantics of the numeral or the silent MANY; although *book* and *books* here will have different meanings for us, as soon as they are fed as arguments to MANY these differences are gotten rid of, as pluralization is applied to both (vacuously to the latter, non-vacuously to the former).

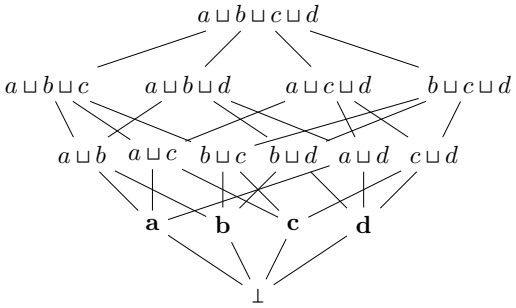


Figure 1: The domain of entities

(26) Nigella hoeft **geen** 20 taarten te bakken.
Nigella must-NPI GEEN 20 cakes to bake
'Nigella doesn't have to bake 20 cakes.'

(27) $[\![$Nigella must bake **geen**$_{=}$ 20 MANY cakes$]\!]$ =
$[\![\textbf{geen}_{=}\ 20]\!](\lambda n.\ \Box\ \exists x[\textbf{*bake}(\textbf{N}, x)\ \&\ \textbf{*cake}(x)\ \&\ \#x = n]) =$
$\neg max\{n \mid \Box\ \exists x[\textbf{*bake}(\textbf{N.}, x)\ \&\ \textbf{*cake}(x)\ \&\ \#x = n]\} = 20$

(28) $[\![$Nigella must bake **geen**$_{\geq}$ 20 MANY cakes$]\!]$ =
$[\![\textbf{geen}_{\geq}\ 20]\!](\lambda n.\ \Box\ \exists x[\textbf{*bake}(\textbf{N}, x)\ \&\ \textbf{*cake}(x)\ \&\ \#x = n]) =$
$\neg\ \Box\ \exists x[\textbf{*bake}(\textbf{N.}, x)\ \&\ \textbf{*cake}\ \&\ \#x = 20]$

The resulting readings are variants of the split-scope reading: it's not the case that Nigella has to bake 20 cakes. The versions differ in that with 'geen$_{=}$', the requirement can be any number other than 20 – higher or lower; with 'geen$_{\geq}$', the requirement is lower than 20. These are indeed the readings available for (26).

3.2 Bare *geen*

We propose that occurrences of *geen* that are not followed by a numeral, as in (29), are derived from the numeral modifier *geen* by semantically incorporating the numeral 'one' (Dutch: *één*). As before, *geen* gives rise to a split scope reading via degree quantifier movement above the modal verb. The split reading is achieved with an 'at least' semantics of *geen* incorporating 'one':

(29) Je hoeft **geen** stropdas te dragen.
You must-NPI GEEN tie to wear.
'You do <u>not</u> have to wear <u>a</u> tie.

(30) $[\![\text{geen}^1_{\geq}]\!] = \lambda P_{\langle dt\rangle}.\neg P(1)$

(31) $[\![$You must wear **geen** tie$]\!]$ =
$[\![\text{geen}^1_{\geq}]\!](\lambda n.\ \Box\ \exists x[\textbf{*wear}(\textbf{u}, x)\ \&\ \textbf{*tie}(x)\ \&\ \#x = n])$
$= \neg\ \Box\ \exists x[\textbf{*wear}(\textbf{u}, x) \& \textbf{*tie}(x) \& \#x = 1]$

(31) expresses the lack of obligation to wear a tie, as desired. Potentially, we could have the second version of *geen* with incorporated 'one', parallel to the prenumeral 'geen$_{=}$':

(32) $[\![\text{geen}^1_{=}]\!] = \lambda P_{\langle dt\rangle}.max\{m|P(m)\} \neq 1$

However, bare *geen* only has the 'at least' reading – that is, (29) only has (31) as a reading. Using (32) in (31) would amount to the lack of obligation to wear exactly one tie. This reading is not attested. Similarly, 'I have **geen** book(s)' with (32) would be a statement that is true in a situation where I have no books or two books, or three books, etc.

We believe there is a very specific reason why bare *geen* does not express the quantificational concept in (32). The reason is that 'geen$^1_=$' denotes a discontinuous fragment of the quantity scale: the complement of 1. This meaning, we suggest, has a disadvantage on a lexicalization path. In particular, we appeal to convexity, or connectedness, of lexical meanings to rule out this lexical entry (cf. Gärdenfors 2004; Jäger 2010; Zwarts and Gärdenfors 2016). A recent version of this idea, due to Chemla (2017), is that whenever the domain of the denotation of a word can be seen as ordered, supporting an in-between relation, it has no gaps. Using Chemla's term, denotations of words are *connected*. A somewhat simplified version of this constraint says that for any three objects o_1, o_2 and o*, if the latter is in between the first two, and o_1 and o_2 belong to the denotation of the word, then o* also belongs to the denotation of the word. The connectedness constraint rules out the possibility of there being a quantifier meaning 'less than 5 or more than 10'. Similarly, one can see this constraint as ruling out 'geen$^1_=$': it has an ordered domain of intervals on the quantity scale (although see Section 4). Closed intervals [0, 0] and [0, 2] have the interval [0, 1] in between (this being a consequence of 1 being in between 0 and 2). 'Geen$^1_=$' assigns 'True' to [0, 0] and [0, 2] but not to [0, 1], therefore, we take 'geen$^1_=$' to have a gapped denotation in the sense described above, and it is therefore predicted to have a lexicalization disadvantage.

An indirect indication of this restriction comes from *geen* in combination with overt numeral 'one': *geen één* ('**geen** one'). With normal prosody, this combination does get the discontinuous interpretation that is unavailable for bare *geen*. However, when 'one' is deaccented and forms a prosodic unit with *geen*, the 'exactly'-interpretation becomes unavailable. This suggests that the lexicalization process indeed avoids gapped denotations, and 'geen$^1_=$' might be one of them.

(33) Ze heeft geen één boek gelezen, maar twee.
She has GEEN one book read but two
'She didn't read one book, she read two'.

(34) Ze heeft geen-één boek gelezen, #maar twee.
She has GEEN-one book read but two
'She didn't read one book, she read two'.

4 Extensions

Other uses of *geen* / *kein* — Extensions of our analysis cover two further uses of *geen* / *kein*: i) combinations with mass nouns like in (35); ii) seemingly non-quantificational cases like (36) (both examples from Dutch):

(35) Nigella heeft geen soep gemaakt.
N. has no soup made.
'Nigella didn't make soup'

(36) Hij is geen genie
He is GEEN genius
'He is not a genius'

We analyze both cases by moving from a discrete cardinality scale as the domain of *geen* to a

dense scale. Both examples above involve instances of '$\text{geen}^{1}_{\geq}$', but in the case of combinations with mass nouns, '$\text{geen}^{1}_{\geq}$' makes reference not to number '1' but rather to its correlate on a dense scale – the lowest non-zero degree on the dense quantity scale (1 being its correlate on the discrete quantity scale). (35) then states the lack of such non-zero degree that would make the statement 'Nigella made that much soup' true.

In the case of (36), the domain of '$\text{geen}^{1}_{\geq}$' is again a dense domain, but not a numeric one – instead, it consists of degrees of genius. Non-numeric '$\text{geen}^{1}_{\geq}$' negates that the lowest non-zero degree on the relevant scale holds of the subject. Crucially, like the cases discussed above, such non-quantificational negative indefinites split in Dutch/German, but not in English.

(37) Jan hoeft geen genie te zijn.
Jan needs no genius to be.
'Jan doesn't need to be a genius.'

(38) Jan has to be no genius. (*no split reading*)

We conclude that the meaning of *geen* / *kein* is more general than the discreet cardinality meaning that we developed in Section 3 to cover the basic readings. However, the corresponding extensions are relatively straightforward, as formulated above.

Focus sensitivity — The present account makes similar predictions to the theory of split scope in Blok (2018). Blok argues that the unlimited ability to give rise to split scope readings is a property of focus-sensitive operators. Split readings arise when these operators move over another scope-bearing element, leaving behind their DP complement. Crosslinguistic data provide evidence for what we might call the *focus sensitivity generalization*: whenever an expression is focus-sensitive, it will give rise to split scope readings across the board. This includes expressions we consider degree expressions in this paper: *at least*, *at most*, and negative indefinites in Dutch, German, Frisian, and Icelandic. It excludes negative indefinites in English, Swedish, Danish, and Norwegian. Thus, the empirical picture that ensues is very similar. In addition, the numeral modifier generalisation mentioned in section 2.4 of this paper can be subsumed by the focus-sensitivity generalization. As mentioned there, there is a correlation between the ability to modify numerals and the unlimited ability to create split readings. Blok argues that focus-sensitivity is at the root of this correlation: focus-sensitive expressions yield split readings and are also known for their ability to modify a wide range of different types of expressions, including numerals. One area where the present account differs from Blok (2018) is in the predictions regarding comparative numeral modifiers such as *fewer than* and the Heim-Kennedy generalization. See Blok (2018) for a discussion of these matters and for reasons why the predictions of the two accounts may actually not be as different as they seem.

5 Discussion

We argued that split scope as observed with *geen*-type indefinites is essentially a degree phenomenon. Our analysis of *geen* makes it a degree quantifier, therefore split scope items form a natural class – degree quantifiers. English *no* is not a degree quantifier, as seen in its inability to combine with numerals – unlike *geen*. The mechanism of split scope is that of degree quantifier raising.

We believe that this analysis has an advantage over other existing analyses of split scope with *geen*-type expressions, none of which systematically account for the discrepancy between *geen* and degree quantifiers on the one hand and *no* on the other hand. Existing analyses of split scope can be divided into a class of decompositional analyses and a class of higher-

type analyses. The former treat *geen* as semantically and/or syntactically complex, multiple components being spelled out as one word (Rullmann, 1995), or, alternatively, as a positive indefinite that needs to be licensed by sentential negation (Penka and Zeijlstra, 2005; Penka, 2011). Higher-type analyses come in two flavours: quantification over properties (de Swart, 2000) and quantification over choice functions (Abels and Martí, 2010). According to the former, split scope readings arise when a negative DP QRs, and then a type lifting operation takes place, so that the quantifier quantifies over properties rather than over individuals. According to the latter, natural language determiners are uniformly quantifiers over choice functions. In the case of split scope, after the negative DP QRs, selective deletion takes place: in *no tie*, *tie* is deleted upstairs and *no* is deleted downstairs. Under all of these views, parallels between *geen*-type indefinites come as a mere coincidence, and the difference between *geen* and *no* remains unaccounted for – unlike under the view we propose here.

This said, there are two issues that we have left open. First of all, we have not said anything about cases when split readings of *geen* occur with quantifiers over individuals under hat contour, breaking the Heim-Kennedy generalization, as in (39) from German.[3] We do not have an analysis of such cases and leave them for future work.

(39) /JEDER Arzt hat KEIN\ Auto
every doctor has no car
'Not every doctor has a car'

Similarly, we do not give an analysis of cases when English *no* does give rise to split scope, as was the case for (4), repeated here as (40).

(40) The company need fire **no** employees.
'It is <u>not</u> the case that the co. is obligated to fire <u>an</u> employee.' $\neg > \Box > \exists$

All we say about these examples is that the mechanism must be different from what we suggest for *geen* and other degree quantifiers.

Rather ironically, our analysis suggests then that the *true* split scope puzzle is found not in languages like Dutch or German, where split scope examples involve a rather humdrum form of degree quantifier raising, but rather in languages like English, where in a very restricted set of contexts non-degree negative indefinites appear to split their scope.

References

Abels, K. and L. Martí (2010). A unified approach to split scope. *Natural language semantics 18*, 435–470.

Blok, D. (2018). Doctoral dissertation, Utrecht University (expected).

Buccola, B. and B. Spector (2016). Modified numerals and maximality. *Linguistics and Philosophy 39*(3), 151–199.

Bylinina, L. and R. Nouwen (2017). On 'zero'. SALT 27.

Chemla, E. (2017). Connecting content and logical words. http://semanticsarchive.net/Archive/WVhYzUwM/Chemla-ConnectWords.pdf.

[3] In fact, de Swart (2000) takes such examples to indicate that scope splitting is not a phenomenon restricted to intensional operators. Note, however, that examples like (i) do not generalise to other nominal quantifiers, like for instance *most*.

de Swart, H. (2000). Scope ambiguities with negative quantifiers. In K. von Heusinger and U. Egli (Eds.), *Reference and anaphoric relations*, pp. 109–132. Dordrecht: Kluwer.

Gärdenfors, P. (2004). *Conceptual spaces: The geometry of thought*. MIT press.

Geurts, B. (1996). On 'no'. *Journal of Semantics 13*, 67–86.

Hackl, M. (2000). *Comparative quantifiers*. Ph. D. thesis, MIT.

Heim, I. (2000). Degree operators and scope. In *Proceedings of SALT 10*, Ithaca, NY. CLC Publications.

Iatridou, S. and I. Sichel (2011). Negative DPs, A-movement, and scope diminishment. *Linguistic Inquiry 42*, 595–629.

Jacobs, J. (1980). Lexical decomposition in Montague Grammar. *Theoretical lingusitics 7*, 121–136.

Jäger, G. (2010). Natural color categories are convex sets. In *Logic, language and meaning*, pp. 11–20. Springer.

Kennedy, C. (1997). *Projecting the adjective*. Ph. D. thesis, UCSC.

Kennedy, C. (2015). A de-Fregean semantics (and neo-Gricean pragmatics) for modified and unmodified numerals. *Semantics and Pragmatics 8*(10), 1–44.

Kennedy, C. and P. Alrenga (2014). No more shall we part: Quantifiers in English comparatives. *Natural language semantics 22*(1), 1–53.

Kratzer, A. (1995). Scope or pseudoscope? Are there wide-scope indefinites? Ms., University of Massachusetts, Amherst.

Landman, F. (2011). Boolean pragmatics. Ms.

Nouwen, R. (2008). Upper-bounded *no more*: the implicatures of negative comparison. *Natural Language Semantics 16*(4), 271–295.

Nouwen, R. (2010). Two kinds of modified numerals. *Semantics and Pragmatics 3*(3), 1–41.

Nouwen, R. and J. Dotlačil (2017). The scope of nominal quantifiers in comparative clauses. To appear in Semantics & Pragmatics.

Penka, D. (2011). *Negative indefinites*. Oxford, UK: Oxford University Press.

Penka, D. and H. Zeijlstra (2005). Negative indefinites in Dutch and German. Ms., University of Tuebingen.

Potts, C. (2000). When even 'no's Neg is splitsville. In S. Chung, J. McCloskey, and N. Sanders (Eds.), *Jorge Hankamer webfest*. Santa Cruz, CA: Linguistics Research Center.

Rullmann, H. (1995). Geen einheid. *Tabu 25*, 194–197.

von Fintel, K. and S. Iatridou (2007). Anatomy of a modal construction. *Linguistic Inquiry 38*, 445–483.

Zwarts, J. and P. Gärdenfors (2016). Locative and directional prepositions in conceptual spaces: The role of polar convexity. *Journal of Logic, Language and Information 25*(1), 109–138.

Ignorance Implicatures and Non-doxastic Attitude Verbs*

Kyle Blumberg

New York University
khb251@nyu.edu

Abstract

This paper is about conjunctions and disjunctions in the scope of non-doxastic attitude verbs. These constructions generate a certain type of ignorance implicature. I argue that the best way to account for these implicatures is by appealing to a notion of contextual redundancy (Schlenker, 2008; Fox, 2008; Mayr and Romoli, 2016). This pragmatic approach to ignorance implicatures is contrasted with a semantic account of disjunctions under 'wonder' that appeals to exhaustification (Roelofsen and Uegaki, 2016). I argue that exhaustification-based theories cannot handle embedded conjunctions, so a pragmatic account of ignorance implicatures is superior.

1 Introduction

This paper is about conjunctions and disjunctions in the scope of non-doxastic attitude verbs. To see what is at issue, consider the following scenarios and reports that follow them (the embedded question in (2b) is a disjunctive polar question rather than an alternative question)[1]:

Visitors: On Friday, Bill gets a letter from his friends Alice and Ted saying that they will visit Bill on Sunday if they find enough free time. On Saturday, Bill gets a message from Alice saying that she won't be able to manage a visit — the message is silent about the prospects of Ted visiting. On Sunday, Bill hears a knock on the door and rushes to open it. Before Bill answers, I utter:

(1) a. Bill hopes that Ted is at the door.
b. ?? Bill hopes that Alice or Ted is at the door.

(2) a. Bill wonders whether Ted is at the door.
b. ?? Bill wonders whether-or-not Alice or Ted is at the door.

Dessert: Bill is having a dinner party and each guest brought something to eat. Bill's favorite desserts are apple pie and cherry pie. Bill sees that Mary brought apple pie, but he doesn't yet know what Chris brought. I utter:

(3) a. Bill hopes that Chris brought cherry pie.
b. ?? Bill hopes that Mary brought apple pie and Chris brought cherry pie.

(4) a. Bill wonders whether Chris brought cherry pie.

*For helpful feedback and discussions I'd like to thank Chris Barker, Cian Dorr, Ben Holguín, Jim Pryor, and four anonymous reviewers.

[1]Disjunctive polar questions are distinguished from alternative questions by their intonation contours (Biezma and Rawlins, 2012), as well as the fact that alternative questions, but not disjunctive polar questions, presuppose that exactly one of the relevant disjuncts hold. I follow others in using 'whether-or-not' for disjunctive polar questions. See §5 for further discussion.

b. ?? Bill wonders whether Mary brought apple pie and Chris brought cherry pie.

While (1a)-(4a) are acceptable in their respective contexts, (1b)-(4b) are not. Intuitively, what seems to be required for (1b)-(2b) to be acceptable is that it is compatible with Bill's knowledge that Alice is at the door; and what seems to be required for (3b)-(4b) to be acceptable is that it is compatible with Bill's knowledge that Mary did not bring apple pie. That is, Bill cannot know that Alice will not be coming, and he cannot know that Mary brought apple pie. Let us call these inferences *ignorance implicatures*.

I argue that the best way to account for ignorance implicatures is by appealing to a notion of *contextual redundancy*. In short, (1b)-(4b) are infelicitous because they have constituents that are redundant in context: the propositions that they express could have been expressed by syntactically simpler sentences, namely (1a)-(4a). This pragmatic approach to ignorance implicatures stands in contrast to a recent semantic account of ignorance implicatures involving disjunctions under 'wonder' developed by Roelofsen and Uegaki (2016) (henceforth 'R&U'). I argue that R&U's account makes problematic predictions when conjunctions are embedded under 'wonder', as in (4b). Thus, the pragmatic, redundancy-theoretic account is superior.[2]

2 Redundancy and Ignorance Implicatures

2.1 Redundancy

Consider the following scenarios and reports that follow them:

Wimbledon: We are watching the men's Wimbledon semi-final. Unfortunately, we all see Federer lose to Nadal in five sets. Then I utter:

(5) ?? Federer won or Nadal will win the final.

Holiday: A group of us are discussing our holiday plans. I ask Ted where he intends to spend the summer. He tells the group: 'I'm going to Costa Rica'. Then Ben utters:

(6) ?? Ted is going to Costa Rica and it is going to be very humid there.

Neither (5) nor (6) are felicitous in their respective contexts. Intuitively, this is explained by the fact that both have parts that are trivial or redundant in the relevant scenarios (in (5) this is the first disjunct, and in (6) this is the first conjunct). That is, the content communicated by (5) and (6) could have been communicated by simpler sentences. If we suppose that more economical expressions are preferred to more complex ones, the unacceptability of (5) and (6) can be accounted for.[3] I maintain that a similar account of the infelicity of (1b)-(4b) can be given: these reports are problematic because their content could have been expressed by simpler sentences in context.

A theory that explains why (5) and (6) are redundant in their respective contexts is a *theory of redundancy*. A rather simple theory of redundancy accounts for (5) and (6), as well as (1b)-(4b):

[2] As a reviewer points out, ignorance implicatures also arise with disjunctions embedded under doxastics, e.g. 'Bill believes that Alice or Ted is at the door' is infelicitous when it is common knowledge that Bill believes Alice is not at the door. The account developed here can handle these cases as well. However, we focus on non-doxastics since, unlike both 'hope' and 'wonder', conjunctions under 'believe' do not give rise to ignorance implicatures.

[3] This is only to say that this is *one* way to account for their infelicity, there could be other explanations as well.

(7) Redundancy 1: (to be revised)
   a. ϕ cannot be used in context C if ϕ is contextually equivalent[4] to ψ, and ψ is a simplification of ϕ.
   b. ψ is a simplification of ϕ if ψ can be derived from ϕ by replacing nodes in ϕ with their subconstituents

To illustrate, (5) is contextually equivalent to 'Nadal will win the final' in $Wimbledon$, since every world in the context set is one in which Fed lost the match. Since 'Nadal will win the final' is a simplification of (5) (by (7b)), (5) is predicted to be unacceptable (by (7a)). Similarly, (6) is contextually equivalent to 'It is going to be very humid in Costa Rica' in *Holiday*, since every world in the context set is one in which Ted is going to Costa Rica. Since 'It is going to be very humid in Costa Rica' is a simplification of (6), (6) is predicted to be unacceptable.

2.2 Some attitude semantics

2.2.1 'hope'

(1a)-(4a) are simplifications of (1b)-(4b), respectively. So, if we can show contextual equivalence for each pair then we would have an explanation for the (b) member's infelicity. In order to show contextual equivalence we need to have a semantics for 'hope' and 'wonder' on the table. For 'hope' let us assume a simplified "ideal worlds" analysis (von Fintel, 1999). This acccount employs a notion of an "ideal" set of worlds with respect to a subject's desires: a set of worlds compatible with everything that S desires in w (denoted by $\text{Bul}_{w,S}$). On this approach, 'S hopes that p' is defined at w iff S does not believe p, S does not believe $\neg p$, and S's hopes are constrained by S's beliefs ($\text{Bul}_{w,S} \subseteq \text{Dox}_{w,S}$).[5] If defined, the report is true iff all of S's desire worlds are p-worlds. A bit more formally:

(8) Semantics for 'hope'
   a. 'S hopes that p' is defined at w iff (i) $\text{Dox}_{w,S} \cap p \neq \emptyset$, (ii) $\text{Dox}_{w,S} - p \neq \emptyset$, (iii) $\text{Bul}_{w,S} \subseteq \text{Dox}_{w,S}$
   b. If defined, 'S hopes that p' is true at w iff $\text{Bul}_{w,S} \subseteq p$

It is straightforward, but tedious, to show that (1a)-(1b) and (3a)-(3b) are contextually equivalent on this semantics for 'hope'.[6] Thus, both (1b) and (3b) are predicted to be unac-

[4]Sentences ϕ and ψ are contextually equivalent with respect to context C iff $\{w \in C : [\![\phi]\!](w) = 1\} = \{w \in C : [\![\psi]\!](w) = 1\}$ Singh (2011).

[5]As Heim (1992) points out, 'I hope to teach Tuesdays and Thursdays next semester' can be true even when there are worlds compatible with everything that I desire in which I don't teach at all. Instead, hope reports only make a claim about the relative desirability of the worlds *compatible with the subject's beliefs*. (As Heim (1992) notes, the relevant constraint isn't quite the subject's belief worlds, but as far as I can tell this subtlety shouldn't impact our argument.)

[6]Let us call the context of $Visitors$ V. Take an arbitrary $w \in V$. Suppose that (1a) is undefined at w. Then at least one of (i)-(iii) in (8) fail with respect to (1a). If (iii) fails then clearly (1b) is also undefined at w. If (i) fails, then at w it is doxastically impossible for Bill that Ted is at the door. Since it is doxastically impossible for Bill that Alice is at the door, it follows that (1b) is undefined at w. If (ii) fails, then at w it is doxastically necessary for Bill that Ted is at the door. It follows that it is doxastically necessary that Ted or Alice is at the door, hence (1b) is undefined at w. So, if (1a) is undefined at w, then (1b) is undefined at w. Now suppose that (1a) is defined at w. Then it is doxastically possible but not necessary for Bill that Ted is at the door at w. Since it is doxastically impossible for Bill that Alice is at the door in w, it follows that it is doxastically possible but not necessary for Bill that Ted or Alice is at the door. Furthermore, if (1a) is defined at w then condition (iii) of (8) is satisfied. Thus, if (1a) is defined at w, (1b) is defined at w. Now suppose that (1a) is true at w. Then all of the worlds compatible with what Bill desires are worlds in which Ted is at the door. Hence, all of

ceptable given Redundancy 1.[7] More generally, if it is common knowledge that S believes p is false, 'S hopes that p or q' will be contextually equivalent to 'S hopes that q'. Thus, by Redundancy 1 the report will be unacceptable. Similarly, if it is common knowledge that S believes p is true, 'S hopes that p and q' will be contextually equivalent to 'S hopes that q'. Thus, by Redundancy 1 the report will be unacceptable.[8]

2.2.2 'wonder'

I will assume the semantics for 'wonder' developed by Ciardelli and Roelofsen (2015). Their theory is set in the framework of inquisitive epistemic logic, which combines notions from standard epistemic logic and inquisitive semantics. In epistemic logic, an *information state* is modeled as a set of possible worlds—those worlds that are compatible with the information available in the state. In inquisitive semantics, the basic propositional object is an issue I: a non-empty set of information states that is closed under subsets, i.e. if $s \in I$ and $s' \subset s$ then $s' \in I$. The maximal elements of I are called the *alternatives* of I. The *meaning of a sentence*, whether declarative or interrogative, is the issue that it expresses. For example, $[\![$whether Ted is at the door$]\!] = \{s \mid \forall w \in s :$ Ted is at the door in $w\} \cup \{s \mid \forall w \in s :$ Ted is not at the door in $w\}$. An information state s *settles* an issue I iff $s \in I$. For instance, if Ted is at the door at the actual world $w_@$, then $\{w_@\}$ settles the issue of whether Ted is at the door.

Each agent α is assigned an *inquisitive state* at a world w denoted as $\Sigma_\alpha(w)$: a set of information states such that each information state settles all the issues that α entertains at w. For instance, if at w Bill entertains the issue of whether Ted is at the door, then every $s \in \Sigma_{\text{Bill}}(w)$ settles that issue. Intuitively, Σ_α tells us 'where the agent wants to get to' in terms of inquiry; how they would like their information state to be in the future, and which issues they want to see settled. Like issues, inquisitive states are assumed to be non-empty and closed under subsets. Moreover, it is assumed that $\Sigma_\alpha(w)$ forms a *cover* of α's information state at w, denoted as $\sigma_\alpha(w)$. That is, $\bigcup \Sigma_\alpha(w) = \sigma_\alpha(w)$.

In this system, α *knows* an issue I at w when $\sigma_\alpha(w) \in I$. α *entertains* an issue I when $\Sigma_\alpha(w) \subseteq I$ (all of the information states that α would like to get to are ones where I is settled). The 'wonder' modality, denoted W, is given in terms of these notions and has the following truth conditions: $w \models W_\alpha\phi$ iff$_d$ $\sigma_\alpha(w) \not\in [\![\phi]\!]$ and $\Sigma_\alpha(w) \subseteq [\![\phi]\!]$. Finally, the semantics for 'wonder' is given in terms of this modality:

(9) Semantics for 'wonder'
'S wonders ϕ' is true at w iff $w \models W_S\phi$ (iff $\sigma_S(w) \not\in [\![\phi]\!]$ and $\Sigma_S(w) \subseteq [\![\phi]\!]$)

In other words, S wonders about an issue when they do not know it, but would like to see it settled, i.e. they entertain it. It is easy to check that (9) makes 'wonder' non-monotonic (since the underlying 'wonder' modality is non-monotonic).[9]

---

the worlds compatible with what Bill desires are worlds in which Ted or Alice is at the door. So, if (1a) is true at w, (1b) is true at w. Finally, suppose that (1a) is false at w. Then it is not the case that all of the worlds compatible with what Bill desires are worlds in which Ted is at the door. Since Bill's desire worlds are a subset of his belief worlds, it follows that it is not the case that all of the worlds compatible with what Bill desires are worlds in which Ted or Alice is at the door. Thus, if (1a) is false at w, (1b) is false at w. Hence, (1a) and (1b) are contextually equivalent with respect to V. The other case is similar.

[7] The same result obtains if a "similarity" semantics for 'hope' is adopted (Heim, 1992).

[8] Note that the "Presupposed Ignorance Principle" of Spector and Sudo (2017) does not predict that either (1b) or (3b) should be unacceptable in their respective contexts, since the negative and positive presuppositions of 'hope' create a non-monotonic environment. See (Spector and Sudo, 2017) for further discussion.

[9] In this framework, for issues I, G: $I \models G$ iff $I \subseteq G$. See (Ciardelli et al., 2016) for more on the logic of issues.

Given (9), it is straightforward to show that (2a)-(2b) and (4a)-(4b) are contextually equivalent.[10] Similar cases involving alternative, rather than polar questions can also be handled (but see §5 for further discussion).

To be clear, we have explained why, e.g. 'S hopes that p or q' is unacceptable when, e.g. it is common knowledge that 'S knows $\neg p$' is true. However, what might be more naturally called an "ignorance implicature" is the following phenomenon: 'S hopes that p or q' uttered *out of the blue* suggests that (the speaker thinks that) 'S knows $\neg p$' is $false$. The account presented here predicts something weaker; namely that such an utterance will merely suggest that it is not common knowledge that 'S knows $\neg p$' is true. That is, what is predicted is $\neg$CK(S knows $\neg p$), but what is required is CK($\neg$(S knows $\neg p$)). It is plausible that the strengthened result is obtained by an "epistemic step" similar to those that have been proposed for inferences involving scalar implicatures, e.g. (Sauerland, 2004), and presuppositions, e.g. (Chemla, 2007). We leave the development of an account of such auxiliary pragmatic reasoning for future work.

3 A refinement

In this section, we refine the account of redundancy introduced above by considering some data that has recently been discussed by Rostworowski (forthcoming). In the course of trying to defend the Russellian analysis of definite descriptions, Rostworowski considers reports such as the following:

(10) a. Bill hopes that the dictator is dead and was assassinated.
b. Bill wonders whether the dictator is dead and was assassinated.

(11) a. Bill hopes that Mary is pregnant and expecting a daughter.
b. Bill wonders whether Mary is pregnant and expecting a daughter.

These reports raise two issues. First, a report such as (10a) is unacceptable if Bill already knows that the dictator is dead. Redundancy 1 can explain this: (10a) and 'Bill hopes that the dictator was assassinated' are contextually equivalent in any context in which Bill knows that the dictator is dead, so (10a) is ruled infelicitous. There are, however, contexts in which (10a) is acceptable, e.g. when Bill has no idea about the health of the dictator. But Redundancy 1 predicts that (10a) will *always* be infelicitous. This is because 'The dictator was assassinated' entails 'The dictator is dead'. So, (10a) and 'Bill hopes that the president was assassinated' are contextually equivalent in *any* context. What is needed, then, is an account that predicts that (10a) is problematic only in contexts where Bill knows that the dictator is dead.

Intuitively, the reason that (10a) can be acceptable is that the second conjunct adds information to the first conjunct: once we have processed the first conjunct it is compatible with what we know that the second conjunct is false. What needs to be done is somehow incorporate the fact that we process sentences in linear order into the redundancy conditions. Thankfully, this has already been done for us by Mayr and Romoli (2016) (following Fox (2008), who in

[10] Take an arbitrary $w \in V$. Suppose that (2a) is true in w. Then $\sigma_{\text{Bill}}(w) \notin [\![\text{whether Ted is at the door}]\!] = \{s \mid \forall w \in s : \text{Ted is at the door in } w\} \cup \{s \mid \forall w \in s : \text{Ted is not at the door in } w\}$. Also, $\Sigma_{\text{Bill}}(w) \subseteq [\![\text{whether Ted is at the door}]\!]$. $[\![\text{Whether-or-not Alice or Ted is at the door}]\!] = \{s \mid \forall w \in s : \text{Alice or Ted is at the door in } w\} \cup \{s \mid \forall w \in s : \text{neither Alice nor Ted is at the door in } w\}$. $\sigma_{\text{Bill}}(w) \cap \{w \mid \text{Alice is at the door in } w\} = \emptyset$ (by assumption). It follows that $\sigma_{\text{Bill}}(w) \notin \{s \mid \forall w \in s : \text{Alice or Ted is at the door in } w\}$, and that $\sigma_{\text{Bill}}(w) \notin \{s \mid \forall w \in s : \text{neither Alice nor Ted is at the door in } w\}$. Thus, $\sigma_{\text{Bill}}(w) \notin [\![\text{whether-or-not Alice or Ted is at the door}]\!]$. Given that $\Sigma_{\text{Bill}}(w)$ covers $\sigma_{\text{Bill}}(w)$, it also follows that $\Sigma_{\text{Bill}}(w) \subseteq [\![\text{whether-or-not Alice or Ted is at the door}]\!]$. Hence, (2b) is true in w. The other direction is similar (as is the other case).

turn follows Schlenker (2008)). The result is a more complex redundancy condition that allows us to talk about *parts* or constituents of sentences being redundant:

(12) Redundancy 2
Incremental non-redundancy condition: ϕ cannot be used in context C if any part ψ of ϕ is incrementally redundant in ϕ given C.

a. Incremental redundancy:
   i. ψ is incrementally redundant in ϕ given a context C if it is globally redundant in all ϕ', where ϕ' is a possible continuation of ϕ at point ψ.
   ii. ϕ' is a possible continuation of ϕ at point ψ iff it is like ϕ in its structure and number of constituents, but the constituents pronounced after ψ are possibly different.

b. Global redundancy:
   i. ψ is globally redundant in ϕ given a context C if ϕ is contextually equivalent to ϕ', where ϕ' is a simplification of ϕ without ψ.
   ii. ψ is a simplification of ϕ if ψ can be derived from ϕ by replacing nodes in ϕ with their subconstituents.

Redundancy 2 handles Rostworowski's reports. First, the ignorance implicature of, e.g. (10a) is predicted, since the first conjunct in the complement is incrementally redundant in any context where it has been established that Bill knows that the dictator is dead (the first conjunct is globally redundant in any possible continuation of (10a) at the point of the first conjunct). Moreover, Redundancy 2 does *not* predict that (10a) is always infelicitous. In contexts where Bill does not know that the dictator is dead, there are continuations of (10a) at the point of the first conjunct that are not globally redundant, e.g. 'Bill hopes that the dictator is dead and Mary is happy'.

4 Roelofsen and Uegaki's (2016) account

R&U take as their point of departure Ciardelli and Roelofsen's (2015) semantics for 'wonder' and try to develop an account that captures the ignorance implicatures of disjunctions embedded under this verb. R&U enrich Ciardelli and Roelofsen's semantics with a built-in exhaustivity operator:

(13) R&U's semantics
$\ulcorner \text{wonder } Q \urcorner = \lambda x.\text{EXH}_{\{W_x(\ulcorner Q' \urcorner) | Q' \lesssim Q\}}\ W_x(\ulcorner Q \urcorner)$[11]

(13) can account for the ignorance implicatures that arise for (2b). On this entry, (2b) is true just in case (14a) is true, (14b) is false, and (14c) is false.[12] However, if Bill knows that

[11] The exhaustivity operator takes an expression φ and a set of alternatives $\mathcal{A}$, and 'strengthens' φ by negating every $\psi \in \mathcal{A}$ that is not entailed by φ: $\text{EXH}_{\mathcal{A}}(\varphi) := \varphi \wedge \bigwedge\{\neg\psi |\ \psi \in \mathcal{A} \text{ and } \varphi \not\models \psi\}$ (strictly speaking only the 'innocently excludable' alternatives should be negated, but that complication won't be relevant here). R&U assume that the set of alternatives $\mathcal{A}$ is generated by considering the formal structure of φ, rather than its semantic content. More specifically, $\varphi' \in \mathcal{A}$ with respect to φ just in case $\varphi' \lesssim \varphi$, where $\varphi' \lesssim \varphi$ iff φ' can be obtained from φ by deleting constituents or replacing them with other constituents of the same syntactic category, taken either from the lexicon or from φ itself Katzir (2007).

[12] We leave the complements in English, since it makes the sentences easier to read.

Ann isn't at the door, then the only way for (14a) to be true is for (14c) to be *true*.[13]

(14) a. W_{Bill} (whether-or-not Ann or Ted is at the door)
b. W_{Bill} (whether-or-not Ann is at the door)
c. W_{Bill} (whether-or-not Ted is at the door)

As for conjunctions under 'wonder', R&U's approach does predict that (4b) should be unacceptable. Because W is non-monotonic, both (15b) and (15c) are alternatives for exhaustification for (15a). But if Bill knows that Mary brought apple pie, then (15a) is true only if (15c) is *true*.

(15) a. W_{Bill} (whether Mary brought apple pie and Chris brought cherry pie)
b. W_{Bill} (whether Mary brought apple pie)
c. W_{Bill} (whether Chris brought cherry pie)

Although it captures the relevant ignorance implicature in *Dessert*, overall R&U's semantic approach makes incorrect predictions when conjunctions are embedded under 'wonder'. There are two related problems here. First, the truth-conditions for sentences with conjunctions under 'wonder' seem too strong. It is a consequence of the account that (4b), 'Bill wonders whether Mary brought apple-pie', and 'Bill wonders whether Chris brought cherry pie' cannot all be true together (assuming that the second is false if (15b) is, and that the last is false if (15c) is). But it is quite easy to imagine contexts where all three reports are acceptable, e.g. consider a scenario like *Dessert* where Bill does not know whether Mary brought apple pie. More generally, 'S wonders whether A and B', 'S wonders whether A', and 'S wonders whether B' can all be acceptable in a single context.

Second, R&U's account does not predict ignorance implicatures in all cases. Consider (10b) ('Bill wonders whether the dictator is dead and was assassinated') once again. As discussed above, (10b) is only felicitous if Bill does *not* know that the dictator is dead. However, (16b) is an alternative for exhaustification for (16a):[14]

(16) a. W_{Bill} (whether the dictator is dead and was assassinated)
b. W_{Bill} (whether the dictator is dead)

On R&U's account, (10b) is true only if (16b) is false. (16b) is false just in case either $\sigma_{\text{Bill}}(w) \in [\![\text{whether the dictator is dead}]\!] = \{s \mid \forall w \in s : \text{the dictator is dead in } w\} \cup \{s \mid \forall w \in s : \text{the dictator is not dead in } w\}$ or $\Sigma_{\text{Bill}}(w) \not\subseteq [\![\text{whether the dictator is dead}]\!]$. If Bill knows that the dictator is dead, then $\sigma_{\text{Bill}}(w) \in \{s \mid \forall w \in s : \text{the dictator is dead in } w\} \subseteq [\![\text{whether the dictator is dead}]\!]$. Thus, R&U's account does *not* predict that (10b) is unacceptable when Bill knows that the dictator is dead.[15]

[13] If Bill knows that Ann isn't at the door in w, then $\sigma_{\text{Bill}}(w) \notin [\![\text{whether-or-not Ann or Ted is at the door}]\!]$ only if $\sigma_{\text{Bill}}(w) \notin [\![\text{whether-or-not Ted is at the door}]\!]$. Also, $\Sigma_{\text{Bill}}(w) \subseteq [\![\text{whether-or-not Ann or Ted is at the door}]\!]$ only if $\Sigma_{\text{Bill}}(w) \subseteq [\![$ whether-or-not Ted is at the door$]\!]$, since $\Sigma_{\text{Bill}}(w)$ covers $\sigma_{\text{Bill}}(w)$.

[14] In inquisitive semantics, [[whether the dictator is dead and was assassinated]] = [[whether the dictator was assassinated]]. So, 'W_{Bill} (whether the dictator was assassinated)' is not an alternative for exhaustification for (16a), since the latter entails the former.

[15] Since 'hope' carries presuppositions, an analogue of the exhaustification entry for this verb presents various options depending on how the exhaustification operator is defined. Spector and Sudo (2017) consider some of these alternatives. Overall, these alternatives struggle with embedded conjunctions. Briefly, if EXH_1 is used then it is predicted that (3b) should *always* be infelicitous. Alternatively, EXH_2 does not generate any alternatives at all for (3b) assuming an "ideal worlds" semantics, so cannot account for its ignorance implicatures. If a "similarity" semantics is adopted then EXH_2 raises problems similar to those raised by R&U's account, namely the truth conditions of (3b) are too strict and the ignorance implicatures of (10a) are not accounted for.

5 Further issues

Here we consider some concerns that have been raised about the pragmatic account developed above, as well as pragmatic treatments of ignorance implicatures more generally. First, R&U point out that ignorance implicatures involving *alternative* questions under 'wonder' seem to be *local* in the sense that they take scope below operators, e.g. quantifiers, that are syntactically above this verb:

Crime: There is a crime with three suspects, Ann, Bill, and Carol. There are five detectives investigating the case; one has already ruled out Carol but is still wondering whether it was Ann or Bill. The others don't know anything yet. I say:

(17) Exactly four detectives are wondering whether it was Ann, Bill, or Carol.

As R&U comment, (17) is acceptable in context. However, it is $false$ on (9) since *all five* detectives are such that (i) they do not know whether it was Ann, Bill or Carol, and (ii) every information state they want to be in resolves the issue of whether it was Ann, Bill or Carol. So, the pragmatic, redundancy-theoretic approach cannot capture our judgments, although R&U's theory can.

However, the empirical picture here is rather complex. For one thing, embedded *disjunctive polar* questions do *not* always seem to pattern the way of (17), nor do embedded disjunctions under 'hope':

Cake: Bill and Alice run a birthday cake delivery service. Five of my friends are waiting for a delivery for my surprise party. Everyone knows that either Alice or Bill will make the delivery, but Ted is the only one that knows Bill is at home sick. Nobody is sure of the exact time of the delivery. The doorbell rings. Consider:

(18) ?? Exactly four people are wondering whether-or-not Bill or Alice is at the door.

(19) ?? Exactly four people hope that Bill or Alice is at the door.

To my ear, (18) is unacceptable in context. This is predicted by (9), since this account makes the report false (*all five* friends are such that (i) they do not know whether-or-not Bill or Alice is at the door, and (ii) every information state they want to be in resolves the issue of whether-or-not Bill or Alice is at the door). However, this report is true on R&U's account, since exhaustification takes place regardless of whether the embedded question is an alternative question or a disjunctive polar question. Similarly, (19) is unacceptable in context. This is predicted on (8), since this account makes the report false (*all five* friends are such that (i) it is doxastically possible but not necessary that Bill or Alice is at the door, and (ii) every desire world is one where Bill or Alice is at the door).

Moreover, the ignorance implicatures generated by embedded conjunctions *also* appear to be local:

Dictator: Five professors heard a rumor that the dictator was killed by a sniper. One of them knows for sure that the dictator is dead but isn't sure how he died. I say:

(20) Exactly four professors are wondering whether the dictator is dead and was assassinated.

(21) Exactly four professors hope that the dictator is dead and was assassinated.

Like (17), (20) is acceptable in context. But just like (17), it is false and thus predicted to be unacceptable on (9). However, it is *also* false on R&U's account, since for *all five* professors x, 'W_x(the dictator was assassinated)' is true. Similarly, (21) is acceptable, yet it is false on (8).

To sum up, the ignorance implicatures of alternative questions embedded under 'wonder' do seem to be local, and thus are not predicted by pragmatic approaches that operate at the utterance level. However, the ignorance implicatures of disjunctive polar questions under 'wonder' as well as disjunctions under 'hope' do not seem to be local, contrary to the predictions of semantic accounts such as R&U's. Furthermore, the ignorance implicatures of embedded conjunctions *do* seem to be local, but this is captured by neither semantic nor pragmatic accounts. Overall, then, the data appears to paint a rather complex picture and does not clearly count in favor of either a pragmatic or semantic approach to ignorance implicatures.

Second, on (9) it makes a semantic difference whether an embedded *alternative* question has exactly two alternatives, or more than two alternatives:
Visitors: Bill knows that either Alice or Ted will visit on Saturday at noon. On Friday, Bill gets a message from Alice saying that she won't be able to manage a visit. At noon on Saturday Bill hears a knock on the door and rushes to open it. Before Bill answers, I utter:

(22) ?? Bill wonders whether Alice or Ted is at the door.

Visitors 2: Bill knows that exactly one of Alice, Chris and Ted will visit Bill on Saturday at noon. On Friday, Bill gets a message from Alice saying that she won't be able to manage a visit. At noon on Saturday Bill hears a knock on the door and rushes to open it. Before Bill answers, I utter:

(23) ?? Bill wonders whether Alice, Chris or Ted is at the door.

Neither (22) nor (23) are acceptable in their respective contexts. However, it is easy to check that (22) is $false$ on (9) while (23) is *true*. Given that the pragmatic account developed here uses (9) as a baseline semantics, it holds that while (22) is false, (23) is merely 'pragmatically unacceptable'. As several anonymous reviewers point out, this does not appear to be a good prediction, since one can respond to (23) with 'That's false, since Bill knows that Alice isn't at the door'. That is, we seem to want to be able to say something *stronger* in response to (23) than what is licensed by the pragmatic account. By contrast, R&U's account predicts that *both* (22) and (23) are false in their respective contexts.

But it is worth noting that it does *not* seem acceptable to respond to (2b) ('Bill wonders whether-or-not Alice or Ted is at the door') with 'That's false, since Bill knows that Alice isn't at the door'. This is not predicted by R&U's account, since (2b) is made false by it. Also, it *is* acceptable to respond to (10b) ('Bill wonders whether the dictator is dead and was assassinated') with 'That's false, since Bill knows that the dictator is dead', but neither the pragmatic approach nor R&U's account predicts this. Once again, the data here does not clearly speak in favor of either a pragmatic or semantic approach to ignorance implicatures.

6 Conclusion

Roelofsen and Uegaki (2016) showed that disjunctions embedded under inquisitive verbs such as 'wonder' generate a certain type of ignorance implicature. I have suggested that a similar sort of ignorance implicature arises from embedded *conjunctions*; moreover, that such implicatures arise for a variety of non-doxastic attitude verbs. On the proposal developed here, ignorance implicatures arising from both disjunctions and conjunctions are handled within the same framework. On this account, these implicatures are fundamentally pragmatic, and can be explained by a suitably sophisticated theory of contextual redundancy. I argued that such an

account is superior to a semantic approach to embedded disjunctions based on exhaustification, since such accounts struggle with embedded conjunctions.

We have made progress on the topic of ignorance implicatures, but it should be clear from our discussion that more work needs to be done. First, as mentioned at the end of §2, the account presented here generates inferences that are often too weak; a strengthening mechanism needs to be developed. Second, the judgments reported in §5 are based on introspection and discussion with only a few native speakers; more work is needed to get a better sense of the empirical landscape. Finally, it is not clear whether the sort of approach to redundancy presented in §3 is ultimately adequate, and perhaps an account that employs local contexts should be used instead (Mayr and Romoli, 2016). However, this would require giving a precise characterization of the local contexts of attitude verbs which, as far as I am aware, has not yet been done.

References

María Biezma and Kyle Rawlins. Responding to alternative and polar questions. *Linguistics and Philosophy*, 35(5):361–406, 2012.

E. Chemla. An epistemic step for anti-presuppositions. *Journal of Semantics*, 25(2):141–173, 2007.

Ivano Ciardelli, Jeroen Groenendijk, and Floris Roelofsen. *Inquisitive semantics*. 2016. Lecture notes.

Ivano A. Ciardelli and Floris Roelofsen. Inquisitive dynamic epistemic logic. *Synthese*, 192(6): 1643–1687, 2015.

Danny Fox. Two short notes on schlenker's theory of presupposition projection. *Theoretical Linguistics*, 34:237, 2008. ISSN 16134060. URL https://www.degruyter.com/view/j/thli.2008.34.issue-3/thli.2008.016/thli.2008.016.xml.

Irene Heim. Presupposition projection and the semantics of attitude verbs. *Journal of Semantics*, 9(3):183–221, 1992.

Roni Katzir. Structurally-defined alternatives. *Linguistics and Philosophy*, 30(6):669–690, 2007.

Clemens Mayr and Jacopo Romoli. A puzzle for theories of redundancy: Exhaustification, incrementality, and the notion of local context. *Semantics and Pragmatics*, 9(7):1–48, November 2016. doi: 10.3765/sp.9.7.

Floris Roelofsen and Wataru Uegaki. The distributive ignorance puzzle. In *Proceedings of Sinn und Bedeutung 21*, 2016.

Wojciech Rostworowski. Descriptions and non-doxastic attitude ascriptions. *Philosophical Studies*, forthcoming.

Uli Sauerland. Scalar implicatures in complex sentences. *Linguistics and Philosophy*, 27(3): 367–391, 2004.

Philippe Schlenker. Be articulate: A pragmatic theory of presupposition projection. *Theoretical Linguistics*, 34(3):157–212, 2008.

Raj Singh. Maximize presupposition! and local contexts. *Natural Language Semantics*, 19 (2):149–168, Jun 2011. ISSN 1572-865X. doi: 10.1007/s11050-010-9066-2. URL https://doi.org/10.1007/s11050-010-9066-2.

Benjamin Spector and Yasutada Sudo. Presupposed ignorance and exhaustification: How scalar implicatures and presuppositions interact. *Linguistics and Philosophy*, 40(5):473–517, 2017.

Kai von Fintel. Npi licensing, strawson entailment, and context dependency. *Journal of Semantics*, 16(2):97–148, 1999. doi: 10.1093/jos/16.2.97. URL http://jos.oxfordjournals.org/content/16/2/97.abstract.

Frege's unification

Rachel Boddy

University of California, Davis

Abstract

The purpose of this paper is to examine Frege's views about the *scientific unification* of logic and arithmetic. In my view, what interpreters have failed to appreciate is that logicism is a project of *unification*, not reduction. The notion of unification, I argue, is especially helpful in clarifying how Frege views the projects of *Grundlagen* and *Grundgesetze*, and the differing role of definitions in these works. This allows us to see that there are two types of definition at play in Frege's logicist works. I further use the notion of unification to offer an interpretation of Frege's notion of fruitful definition, which, I think, helps clarify how the two types of definition relate, and how Frege uses them to ground the unification of logic and arithmetic.

1 Introduction

Frege's logicism is often presented as the thesis that the laws of arithmetic are analytic. According to a particularly influential interpretation, this is an *epistemological* thesis about the nature of arithmetical knowledge. The idea being that Frege's project was to *reduce* arithmetic to logic and, in so doing, to show that arithmetical truths are analytic and, hence, knowledge of them a priori. For this reduction to succeed, Frege required definitions of the core arithmetical concepts, beginning foremost with an explicit definition of number. It appears, however, that Frege's definitions are unable to underwrite the claim that the *Grundgesetze* derivations show that *arithmetical* truths are analytic. Arithmetical truths, presumably, are truths *about numbers*. But Frege's defined concepts do not appear to express *arithmetical content*, and consequently, it is unclear how arithmetical content is preserved in the mathematical "reduction" of arithmetic to logic.[1] This lack of clarity underwrites the basis of Benacerraf's argument that logicism was *not* an epistemological thesis for Frege [Benacerraf(1981)].

Benacerraf's argument has inspired much discussion (and dispute!) in the literature.[2] Problematically, its conclusion appears to be diametrically opposed to several of Frege's own explanations of his project. In my view, what interpreters have failed to appreciate is that logicism is equally a thesis about *logic*, in particular, a thesis about the *expressiveness* of logic. Once we take this into consideration, this raises at once the following two questions: (a) *What content do arithmetical truths express?* and (b) *Is this content derivable within logic?* This second question entails the further question of whether the principles of logic can underwrite the existence of mathematical objects. When logicism is cast as a thesis about *logic*, the central task of Frege's formal derivation of arithmetic within logic is to defend a positive answer to (b), rather than an answer to (a). Furthermore, this derivation would not just show that the truths of arithmetic are reducible to logic, but rather it would show that logic and arithmetic "constitute a unified science" [Frege(1885), 112].

[1] *Grundgesetze*, i.e., Gottlob Frege's *Grundgesetze der Aritmetik* (1903). Throughout this paper, references are to Ebert and Rossberg's translation, i.e., [Frege et al.(2013)Frege, Ebert, and Rossberg]. I shall also use "*Gg*" as an abbreviation. Similarly, I shall abbreviate *Die Grundlagen der Arithmetik* as "*Grundlagen*" or "*Gl*". All references are to Austin's translation, i.e., [Frege and Austin(1980)].

[2] See [Blanchette(1994), Weiner(1984), Tappenden(1995), Jeshion(2001)].

The purpose of this paper is to examine Frege's views about the *unification* of logic and arithmetic. The notion of unification, I argue, is especially helpful in clarifying how Frege views the projects of *Grundlagen* and *Grundgesetze*, and the differing role of definitions in these works (sections 2 and 3). I use the notion of unification to offer an interpretation of Frege's notion of fruitful definition, which, I think, helps clarify how the two notions of definition relate (section 4). Finally, I use the foregoing discussion to address our opening question (section 5): Is the *Fregean* thesis that arithmetical truths are analytic an *epistemological* thesis?

2 Logic and arithmetic as a unified science

Frege clarifies his view of the relationship between logic and arithmetic in the paper "On Formal Theories of Arithmetic" ([Frege(1885)]; hereafter [FTA]), which followed the publication of *Grundlagen*. In [FTA], he presents the logicist thesis as the thesis that logic and arithmetic are a unified science:

> [N]o sharp boundary can be drawn between logic and arithmetic. Considered from a scientific point of view, both together constitute a unified science. [112]

Frege's view is that, from a scientific perspective, there are no relevant distinctions between logic and arithmetic, viz., the domain and the inference rules of arithmetic are part of logic, and arithmetical concepts are definable in (and hence reducible to) logic. The passage continues:

> If we were to allot the most general basic propositions and perhaps also their immediate consequences to logic while we assigned their further development to arithmetic, then this would be like separating a distinct science of axioms from that of geometry.

In *Grundlagen*, Frege argued that arithmetic has the same domain as logic on the grounds that arithmetical principles, like logical laws, govern everything thinkable. This is based on the conception of logic as *universal*, viz., as governing the domain of conceptual thought [Goldfarb(2001)]. On this conception, logical laws express (substantive) truths about any subject matter and these laws are, therefore, fully general.[3] Moreover, since every object of (conceptual) thought can be counted, there appears to be no special domain of arithmetical objects.[4] Frege uses this conclusion to argue that concepts have numbers (i.e., that a statement of number is a claim about a concept) and, more specifically, that numbers are extensions of concepts.[5] In *Grundlagen*, he also underlines the analogy between the relationship of the truths of arithmetic to the truths of logic and the relationship of the theorems of geometry to its axioms: "The truths of arithmetic would then be related to those of logic in much the same way as the theorems of geometry to the axioms" [*Gl*, §17]. These considerations suggest that arithmetic is part of logic (just like the theorems of geometry are part of geometry). If this is correct, Frege says, then the principles of logic underwrite the truths of arithmetic, which means that we can express *arithmetical content* in purely logical terms. It also means that we can show that arithmetical truths are theorems of logic.

[3] Frege's conception of logic can be contrasted with what Goldfarb calls a *schematic* conception of logic, according to which logical laws are schemata, i.e., formulas that are only partially interpreted.

[4] In [FTA] Frege again observes that "just about everything that can be an object of thought" can be counted, from which he draws the same conclusion.

[5] For example, Frege analyzes a statement such as "There are eleven houses on 7th street" as the claim that the concept ¡¡house on 7th street¿¿ is satisfied by eleven objects. It is thus a statement about the cardinality of a concept.

Frege's view is that the formal unification of arithmetic and logic requires two steps: first, the *reduction* of arithmetical concepts by means of definitions, and in addition, the *derivation* of arithmetic within logic. The first step is to show that arithmetical content can be expressed in purely logical terms, whereas the second step is to show that arithmetical truths are theorems of logic. In this line, the project of *Grundlagen* is the *discovery* of the content of arithmetic. The development of arithmetic in *Grundgesetze* builds on these results, however, its project is the *justification* of that content.

***Grundlagen*: Discovery** The *conceptual basis* for Frege's *Grundlagen* definition of Number is the thesis that arithmetic is a branch of logic. Part of his argument for this thesis is that numbers can be identified as extensions of concepts. Given Frege's explicit definition of Number, numbers are extensions of second-level concepts.[6] The definition is intended to show that numbers can be described in purely logical terms. This shows how arithmetical content can be reduced to logical content, i.e., that arithmetical content can be expressed logically.[7]

Extensions of concepts are *logical* objects, i.e., objects whose existence can be inferred on purely logical grounds.[8] The claim that arithmetical propositions can be seen to express truths about *those* objects is based on Frege's arguments for the conception of numbers according to which (a) concepts have numbers, (b) numbers are (abstract) objects and (c) these objects satisfy Hume's Principle (i.e., the principle that equinumerous concepts have the same number).[9] Apart from (c), these theses are based on the presupposition that arithmetical propositions express truths about the domain of logic, and not about some more restricted domain (e.g., the Kantian domain of the intuitable).[10] To show that logic and arithmetic are a unified science, Frege has to further show that "there is no peculiar arithmetical mode of inference that cannot be reduced to the general inference-modes of logic" [Frege(1885), 113]. For example, the principle of induction might be an extra-logical inference rule. If it turns out that the proofs of the basic propositions of arithmetic require an extra-logical mode of inference, then the whole approach is undermined. Frege's view, of course, is that we "have no choice but to acknowledge the purely logical nature of the arithmetical modes of inference" [113].[11] But he also thinks that this requires proof [*Gl*, §1].

The arguments from *Grundlagen* are, therefore, not sufficient to show that arithmetic can be *unified* with logic. Frege has only offered a logicist account of how we come to discover the logical means by which we can express the content of arithmetical truths. But discovering the content expressed by these truths is not sufficient for their *justification*. Indeed, as is well-known, Frege separates the context of discovery from the context of justification: "It not uncommonly happens that we first discover the content of a proposition, and only later give the rigorous proof of it, on other and more difficult lines" [*Gl*, §3]. Note too that this passage is part of Frege's

[6] Here is the definition: "The Number which belongs to the concept F is the extension of the concept "equal to the concept F"" (*Gl*, §68).

[7] Notice that this does not show that the numbers *are* extensions of concepts because it is only reductive over content.

[8] At least, ignoring for the moment the inconsistency of Basic Law V.

[9] To be more precise: what is now known as Hume's Principle is the condition that the number of Fs is equal to the number of Gs if and only if there is a one-to-one correspondence between the Fs and the Gs.

[10] Here (c) encodes the idea that numbers are measures of cardinality and are used for counting. Frege analyzes the notion of cardinal number in terms of the equinumerosity of concepts, such that two concepts are equinumerous if their extensions can be placed in a one-to-one correspondence.

[11] As he explains, "[i]f such a reduction were not possible for a given mode of inference, the question would immediately arise, what conceptual basis we have for taking [the mode of inference] to be correct" [113]. The other options he considers are Kantian "intuition" and observation, and, as he has already argued in *Grundlagen*, neither of these options is tenable.

discussion of the analytic/synthetic and a priori/a posteriori distinctions. For Frege, these distinctions concern the justification of a proposition, rather than its content. Analogously, his logicist view is that an account of the content expressed by arithmetical propositions is not sufficient for their justification, for justification requires proof.

In the conclusion of *Grundlagen*, Frege explains that he hopes "to have made it *plausible*" that the laws of arithmetic are analytic [§87; my emphasis]. Given his notion of analyticity, this means that these laws can be proved using only logical laws and definitions, and thus that arithmetic "becomes simply a development of logic, albeit a derivative one" [§87]. To support this thesis, Frege has offered an explicit definition of cardinal number, and sketches for the proofs that the numbers, as characterized by this definition, have the properties of the natural numbers (see §§70-73). To raise this thesis from plausible to justified, however, requires gap-free derivations of the laws of arithmetic from pure logic. For as long as Frege has not shown that these laws (with their meaning settled as in *Grundlagen*) can be derived within a system of pure logic, it can still be denied, as presumably Kantians would have, that logic can ground the truths of arithmetic.

***Grundgesetze*: Justification** Frege thinks that if arithmetical laws are truths about logical objects (per his analysis in *Grundlagen*) then these laws must be provable in pure logic, and so only by providing such proofs can he vindicate his logicist analysis of Number. As he explains in the foreword of *Grundgesetze*:

> By this act I aim to *confirm* the conception of cardinal number which I set forth in the latter book. The basis of my results is articulated there in §46, namely that a statement of number contains a predication about a concept; and *the exposition here rests upon it.* [*Gg* vol. 1, viii-ix; my emphasis]

At issue is not *what* content sentences of arithmetic express, but rather *whether* these sentences, with their content already settled, are derivable within a system of pure logic.[12]

In *Grundgesetze*, Frege shifts to talk of the "ideal of a rigorous scientific method", according to which proof is constructed in an axiomatic system. This shift corresponds to the shift from "discovering" the content of arithmetical claims (in *Grundlagen*) to that of their justification (in *Grundgesetze*). For Frege, the firmest type of justification is logical proof in an axiomatic system. Such a system, on this view, consists of the complete specification of a language, together with axioms (formulated in that language), inference rules and possibly definitions. Questions about justification, then, can only be treated rigorously in the context of a system, i.e., an entire theory. This also means that whether a proposition is analytic depends on the system in which it is proved.[13]

The task of *Grundgesetze*, then, is explicitly to address the shift from claims about discovery to demonstrations of justification. Its introduction opens thus:

> In my *Grundlagen der Arithmetik* I aimed to make it plausible that arithmetic is a branch of logic... In the present book this is now to be *established by deduction* of the simplest laws of cardinal number *by logical means alone.* [*Gg* vol. 1, 1; my emphasis]

[12] As Frege explains: "Usually, mathematicians are merely concerned with the content of a proposition and that it be proven. Here the novelty is not the content of the proposition, but how its proof is conducted, on what foundation it rests" [*Gg* vol. 1, viii].

[13] See also [Dummett(1991)] for discussion of this point.

Frege assumes that the *Grundgesetze* theorems that are labeled "basic laws of cardinal number " are just concept-script renderings of the basic propositions of arithmetic. According to the *Grundlagen* account, the natural numbers are *cardinal* numbers and thus the basic propositions of arithmetic are the basic laws of cardinal number.[14]

3 Frege's definitions

Prior to *Grundlagen*, Frege briefly discusses definitions in *Begriffschrift*.[15] Definitions, he says, are just stipulations that serve to introduce abbreviations into a language:

> . . . nothing follows from [a definition] that could not be inferred without it. Our sole purpose in introducing such definitions is to bring about an extrinsic simplification by stipulating an abbreviation. They serve besides to emphasize a particular combination of signs in the multitude of possible ones, so that our faculty of representation can get a firmer grasp of it. [Frege(1879), 55]

Frege states definitions as identities, viz., sentences of the form $(a = b)$.[16] Once so stated, a definition immediately turns into an analytic judgment, and, " [s]o far as the derivations that follow are concerned, [it] can therefore be treated like an ordinary judgment" [Frege(1879), 55]. The only *content* expressed by this judgment is trivial: it is an instance of the law of identity $(a = a)$. *From the perspective of logical proof*, definitions are, therefore, redundant.[17] Consequently, Frege requires definitions to be eliminable and non-creative.[18] A definition is *eliminable* when its defined term can be eliminated in favor of its defining phrase in any sentence of the language. Eliminable definitions are such that, in Frege's words, "if the *definiens* occurs in a sentence and we replace it by the *definiendum*, this does not affect the thought at all" [Frege(1914), 208]. A definition is *non-creative* when it is *only* used to stipulate the meaning of a term, and cannot, therefore, help prove any result that could not already be proved prior to its introduction.[19]

Where does this leave the role of definitions in Frege's logicist project? Definitions, it seems, must play two distinct roles. First, in *Grundlagen*, where the goal is to show that arithmetical content can be reduced to logic, Frege needs to provide logical definitions of arithmetical concepts. These definitions must *specify a content* in a way that makes plausible the justification of arithmetic. Second, in *Grundgesetze*, where the goal is justification, Frege needs to

[14] See also [Frege et al.(2013)Frege, Ebert, and Rossberg, vol. 2, 155-6].

[15] *Begriffschrift*, i.e., *Begriffsschrift, a formula language, modeled upon that of arithmetic, for pure thought* (1879). References are to van Heijenoort's translation, i.e., [Frege(1879)].

[16] Frege's view of identity shifts from a metalinguistic view in *Begriffschrift* to an objectual view in *Grundgesetze*. It has been argued that in *Begriffschrift*, Frege uses the "=" sign for coreference. Though for a detailed argument against this interpretation, see [May(2012)].

[17] Frege repeats this on several occasions. For example, in "Logic in Mathematics" he writes that it "appears from this that definition is, after all, quite inessential. In fact considered from a logical point of view it stands out as something wholly inessential and dispensable" [Frege(1914), 208]. And in "Foundations of Geometry: First Series" he writes: "[A definition] is only a means for collecting a manifold content into a brief word or sign, thereby making it easier for us to handle. This and this alone is the use of definitions in mathematics" [Frege and MacGuinness(1984), 274].

[18] Note, however, that his discussion of these requirements only occurs in his later work. Also, though Frege thinks that definitions must be eliminable, he does not use this terminology.

[19] This notion corresponds to the familiar notion of conservativeness, according to which a definition is conservative when it cannot help prove any theorem (not involving the defined term) that would otherwise be unprovable. See [Belnap(1993)] for discussion of the eliminability and conservativeness criteria, and [Boddy(manuscript)] for further discussion of these criteria in Frege's work.

show that the *Grundlagen* definitions can be added as conservative extensions to a system of pure logic, and can be used, subsequently, to derive the basic laws of cardinal number. These two roles suggests that Frege has two notions of definition at play: definitions that arise from (conceptual) analysis, and the *Begriffschrift* notion of definition as mere abbreviation.

***Grundlagen*: Conceptual analysis** It appears, then, that the *Grundlagen* definitions, being the result of Frege's logical-philosophical analysis of arithmetical concepts, are *not* conventions of abbreviation and are, therefore, not expected to be eliminable. These definitions, qua logical definitions of arithmetical terms, must preserve (at least part of) the meaning of their defined terms, terms which are not *new* but already have an established use in mathematical practice. Indeed, these definitions appear to be more akin to what Frege in later work calls "analytic definitions" [Frege(1914)]. An "analytic definition" is the result of a logical analysis of a term "with a long established use" that already has a sense, whose sense is analyzed into a complex expression.[20] Such a "definition" is not an arbitrary stipulation and is, therefore, *not* a definition in the *Begriffschrift* sense at all but "is really to be regarded as on axiom" [210]. Frege contrasts analytic definitions with "definition" *tout court*. These are the familiar type of definitions from *Begriffschrift*. This leaves us asking, however, did Frege regard the *Grundlagen* definitions as proper definitions?

The answer is "yes", but with a caveat. The *Grundlagen* definitions are explanations of the meaning of terms that are intended to be used as eliminable definitions in Frege's *Grundgesetze* proofs. Indeed, the *Grundgesetze* definitions are essentially just the *Grundlagen* definitions [Heck(1993), 269]. The caveat is that definitions are properly speaking only definitions in the context of a particular theory (i.e., what Frege calls a "system"). Frege of course intends to add the *Grundlagen* definitions to his logical system. *For this purpose*, it is only relevant whether these definitions are eliminable and non-creative with respect to *that* system.

Similarly, definitions are *stipulations* about the meaning of new terms *within a system*. The terms introduced via definitions need only be *new to the system*. For example, Frege can introduce a number operator into his logicist system with a stipulative definition but, he says, "if we do this, we must treat it as an entirely new sign which had no sense prior to the definition", and that "[i]n constructing the new system we take no account, logically speaking, of anything in mathematics that existed prior to the new system" [Frege(1914), 211]. The choice for a particular new term is "arbitrary" in that it is only constrained by the rules for the construction of well-formed expressions in the language and the requirement that the term be new (to the language).[21] It does not follow that Frege's choice for his defined terms is arbitrary. The *Grundlagen* definitions, being "analytic" (in the relevant sense), are not at all arbitrary. For these definitions must *specify a content* in a way that allows for the *justification of arithmetic*. This does not preclude these definitions from being *used as* eliminable definitions in *Grundgesetze*, however. For the sake of exposition, I shall call the *Grundlagen* definitions "conceptual definitions," and definitions used to introduce abbreviations (as found in *Begriffschrift* and *Grundgesetze*) "proper definitions".[22]

[20] Thus, Frege explains, we start from "a simple sign with a long established use" and then "give a logical analysis of its sense, obtaining a complex expression which in our opinion has the same sense" [210].

[21] See §§26-28 and 33 of *Grundgesetze* (vol. 1) for the formation rules for names (i.e. terms) in the concept-script, and the rules for constructing definitions. Frege here presents the following "governing principle for definitions: Correctly formed names must always refer to something." This is followed by (what is now known as) the proof of referentiality for the *Grungesetze* names.

[22] To be clear: my use of the term "conceptual definition" differs from Frege's use of the term "analytic definition".

***Grundgesetze*: Gap-free proof** Frege thinks that a successful justification of a reductive analysis of number should result in a definition of number that can be added as a conservative extension to its reducing theory. Moreover, he insists that to *prove the worth of the definition*, it must be shown that it enables the construction of gap-free proofs of the well-known properties of the numbers, as described by the laws of cardinal number. The central task of *Grundgesetze* is the gap-free proof of these laws in Frege's logical system. Hence, Frege returns to the *Begriffschrift* conception of definition. The only difference being that in the *Grundgesetze* definitions, the defined phrase is stipulated to have the same sense and the same reference as the defining phrase.[23] It is not just the justification of the laws of cardinal number that is at issue, Frege also intends to justify his *Grundlagen* definitions. For Frege, the justification of a definition "must be a matter of logic" [*Gl*, ix]. Specifically, the logical justification of a definition consists in the definition satisfying the eliminability and non-creativity requirements.

As used in *Grundgesetze*, the explicit definition of number cannot show that any of the derived theorems are indeed theorems *of arithmetic*. But what it can help show is that logic is sufficiently expressive for the proofs of the laws of cardinal number, such that these laws are already recognized as *arithmetical*.[24] For these proofs show that the derived sentences are grounded on principles of logic only. The definition, being constructed in a language whose primitive vocabulary is purely logical, does not express any non-logical content. It thus shows that no *additional*, non-logical, content is required for the proofs of the laws of cardinal number. In addition, it helps facilitate the recognition of the numbers in the logicist development of arithmetic.

4 *Grundlagen*'s fruitfulness requirement of definitions

As we have seen, there must be an appropriate tie between the two types of definition such that the *Grundgesetze* development of arithmetic can justify the *Grundlagen* conception of cardinal number. It would be a mistake, then, to view the *Grundgesetze* definitions as conceptual definitions, as [Horty(2007)] proposes.[25] Frege thinks that once we present an axiomatic system that is constructed "from the bottom up", like *Grundgesetze*'s system, there is *no need* for conceptual definitions because we can treat the defined terms as entirely new. The task of *Grundgesetze* is the *justification* of (arithmetical) content, not its discovery. Prior to the explicit definition of number, Frege has already concluded, on the basis of his conceptual definition of Number in *Grundlagen*, that numbers can only be logical objects. In *Grundgesetze*, he takes this conception of the numbers for granted.[26]

How, then, can Frege's definitions play their two roles? Frege's answer, in my view, is that to play both roles, definitions must be fruitful. According to Frege, the *Grundlagen* definitions have (logicist) worth only in so far as they allow for the gap-free proof of the laws of cardinal number. If his definition of Number cannot be used to this end, he says, then it "should be rejected as completely worthless" [*Gl*, §70]. In *Grundlagen*, this worth is witnessed by the requirement that definitions be *fruitful*, such that definitions are fruitful when their introduction is necessary for the gap-free proof of the *sentences* in which their defined terms occur.[27] Notice

[23] *Begriffschrift*, of course, predates Frege's bifurcation of meaning into sense and reference.

[24] That is, per *Grundlagen*, these laws express the scientific content of the basic propositions of arithmetic.

[25] According to Horty, Frege's definitions play their two roles *simultaneously*, viz., to introduce abbreviations into the language of logic and to explicate expressions already in use in the language of arithmetic.

[26] See e.g., [*Gg.* vol. 2, 153]. It is of course undermined by Russell's paradox.

[27] In Frege's most explicit formulation of the requirement, he clarifies the notion of fruitful definition as follows: "Those [definitions] that could just as well be omitted and leave no link missing in the chain of our proofs should be rejected as completely worthless" [*Gl*, §70].Frege's notion of fruitful definition has engendered

that, in the case of the definition of number, these sentences are, foremost, the laws of cardinal number.

Frege's view is that definitions should be fruitful because by being fruitful, definitions show that no *additional* content is required for the proofs of the sentences in which their defined terms figure. This shows that these sentences express the content that they are afforded by the definition. That is, fruitful definitions underwrite the proofs of the sentences in which their defined terms occur, and these proofs show that the defined terms are used in these sentences *exactly with their stipulated meaning*. Frege's *Grundgesetze* proofs of the laws of cardinal number confirm that his definition of Number specifies a content in a way that allows for the derivation of arithmetic.[28] It also confirms that the definition identifies *logical* objects as the numbers. Now, only some of the *Grundgesetze* definitions are paired with a conceptual definition (from *Grundlagen*). Whenever there is such a pairing, the fruitfulness of the definition justifies its analytic counterpart, and demonstrates the sense in which the *Grundgesetze* offers a logical development of *arithmetic*.

While Frege initially discusses the fruitfulness requirement in *Grundlagen*, he continues this in "Logic in Mathematics", where he compares unfruitful definitions to stucco-embellishments on buildings and says that, like such embellishments, unfruitful definitions are only "ornamental" and play no role in the actual development of arithmetic. Such presumed definitions are not really definitions, he says, as they do not actually fix the reference of the numerals but are "only included because it is in fact usual to do so" [212]. As in *Grundlagen*, at issue is the *worth* of the definition in underwriting (or justifying) an analysis of its definiendum [Frege(1914)].[29] According to Frege, if the definition of Number is not shown to underwrite the proofs of arithmetical theorems, then it is also not shown that these theorems express truths about the objects identified by the definition as the numbers. In this case, the definition is useless as a conceptual definition and useless as a proper definition.

5 Are ordinary arithmetical truths analytic?

As noted, Benacerraf has argued that the *Fregean* thesis that arithmetical truths are analytic is not an *epistemological* thesis. The basic idea, as expressed in [Benacerraf(1981)], is that if logicism is an *epistemological* project, then Frege's definitions must preserve the "ordinary" meaning of their defined terms because only when the definitions express arithmetical content can the logicist derivation of *Frege's* arithmetic underwrite the thesis that the truths of *ordinary* arithmetic are analytic, and hence yield a priori knowledge.

Benacerraf's contention is that "Frege did *not* expect *even reference* to be preserved by his definitions" [29].[30] Benacerraf observes that Frege allows that there are several ways of reducing arithmetic to logic and, in particular, that he suggests that the numbers need not be

much discussion in the literature. See, e.g., [Benacerraf(1981), Boddy(manuscript), Horty(2007), Shieh(2008), Tappenden(1995), Weiner(1990)].

[28] Note that this content is specified by the definiens of the definition, and so the fruitfulness of Frege's definition of Number depends on whether it can be shown that the referents of the definiens have the well-known properties of numbers.

[29] In [Frege(1914)] Frege does not use the word "fruitful" (*fruchtbar*) though he discusses the same requirement. As [Tappenden(1995)] observes, Frege no longer uses the "fruitful definitions" terminology in his post-1884 writings. The fruitfulness condition is also not among (or implied by) the principles of definition listed in §33 of *Grundgesetze*. However, each of the *Grundgesetze* definitions is fruitful (in the above sense).

[30] This claim should not be confused with Benacerraf's claim, from [Benacerraf(1965)], that in reductionist projects, like Frege's, the definitions of arithmetical terms do not preserve their ordinary meaning (or that meaning does not determine reference) because there are different ways of assigning referents to the mathematical vocabulary.

identified with extensions of concepts, but could have been identified with different referents. This shows, he argues, that Frege's definitions are not expected to preserve the referents of the numerals of ordinary arithmetic.[31] If so, then the sentences derived from those definitions are not expected to express truths of *ordinary* arithmetic either. Hence he concludes that Frege did not intend to show that arithmetic is analytic, and thus yields a priori knowledge.

Benacarraf's argument has inspired much discussion in the literature. Interpreters have focused on the question of whether Frege's project can nonetheless still be taken as an *epistemological* project. Against Benacerraf's interpretation, Frege himself repeatedly says that his concern is with the nature of our knowledge of arithmetic. In this line, [Weiner(1990)] argues that Frege intended to present a theory that was to *replace* ordinary arithmetic because, on his view, the numerals of ordinary arithmetic did not refer prior to his work.[32] There was, therefore, no reference to be preserved by Frege's definitions. Her response to Benacerraf is that Frege's logicist arithmetic can replace ordinary arithmetic because "it has all the applications of arithmetic" [115].

If "ordinary" arithmetic is number theory as "ordinarily understood", where this requires some shared view of the content expressed by the numerals, then Frege does not think that there is an ordinary arithmetic. Hence, there is also not an ordinary arithmetic *to replace*. Indeed, his opening argument in *Grundlagen* is exactly that there is no common understanding of what numbers are. The problem that underlies the *Grundlagen* discussion is exactly that it is not clear *exactly what content* the definitions of arithmetical concepts must preserve. That is, the problem is not that the "ordinary" numerals do not have referents prior to Frege's work, but rather that it is *unclear* what exactly these referents are: most mathematicians have some notion of what numbers are—enough to agree on the truth of arithmetical claims—but this "inkling" is imprecise and unarticulated and, therefore, defective [Frege(1914), 221]. Frege's view is that to correct this shortcoming, he needs to derive the well-known properties of the numbers from his definition of Number. The development of arithmetic in *Grundgesetze* justifies the thesis that principles of pure logic can found arithmetical content. Moreover, what *Grundgesetze* shows is that the justification of Frege's definitions, qua definitions of arithmetical concepts, ultimately depends on their ability to help prove the laws of cardinal number within such a system.[33] If this is correct, then *Frege's* arithmetic is intended to be just *a scientifically founded version of ordinary arithmetic*. So clearly logicism is an epistemological thesis. This was, of course, Dummett's point, and I agree [Dummett(1991)]. However, what interpreters, including Dummett and Benacerraf, have failed to appreciate is that the *Grundgesetze* is primarily a work of *justification*, and together with *Grundlagen*, that logicism is a project of unification, not reduction.

In *Grundlagen*, Frege says that the investigation into the foundations of arithmetic is "a task which is common to mathematics and philosophy" [*Gl*, xviii]. I hope to have shown that he means this quite literally. For Frege, to found arithmetic is nothing less than to undertake the *unification* of these two sciences. This requires two steps: the philosophical (or conceptual) discovery of the *plausibility* of the *reduction* of arithmetical content, as described in *Grundlagen*, and the logical *justification* of this content within a system of pure logic, as

[31] Since, for Frege, the reference of a term is determined by its sense, if definitions need not be reference-preserving, then they also need not be sense-preserving. Here, I shall focus on Benacerraf's argument for the claim Frege's definitions are not expected to preserve the reference of the numerals.

[32] Weiner uses the term "refer" as a technical term, such that a term refers when it is "appropriate for scientific use". On her account, Frege did think that the numerals express some content prior to his work, but that they did not have "scientific reference".

[33] [May and Wehmeier(2016)] make a similar point: "In general, Frege's criterion for the adequacy of definitions is holistic; it depends on what can be proven from the definition. Accordingly the adequacy of the definition of number is shown by the proof from it of the "basic laws of cardinal number"" [3fn.7].

shown in *Grundgesetze*.

References

[Belnap(1993)] Nuel Belnap. On rigorous definitions. *Philosophical studies*, 72(2):115–146, 1993.

[Benacerraf(1965)] Paul Benacerraf. What numbers could not be. *The Philosophical Review*, 74(1): 47–73, 1965.

[Benacerraf(1981)] Paul Benacerraf. Frege: The last logicist. *Midwest Studies in Philosophy*, 6(1): 17–36, 1981.

[Blanchette(1994)] Patricia A Blanchette. Frege's reduction. *History and Philosophy of Logic*, 15(1): 85–103, 1994.

[Boddy(manuscript)] Rachel Boddy. Fruitful definitions, manuscript.

[Dummett(1991)] Michael AE Dummett. *Frege: Philosophy of mathematics*. Harvard University Press, 1991.

[Ebert and Rossberg (eds.)(2016)] Philip A Ebert and Marcus Rossberg (eds.). *Essays on Frege's Basic Laws of Arithmetic*. Oxford University Press, 2016.

[Floyd and Shieh(2001)] Juliet Floyd and Sanford Shieh. *Future pasts: the analytic tradition in twentieth-century philosophy*. Oxford University Press, 2001.

[Frege(1879)] Gottlob Frege. Begriffsschrift, eine der arithmetischen nachgebildete formelsprache des reinen denkens. halle. 1879. translated in van heijenoort j. begriffsschrift, a formula language, modeled upon that of arithmetic, for pure thought. *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, pages 3–82, 1879.

[Frege(1885)] Gottlob Frege. On formal theories of arithmetic. *Collected papers*, pages 203–250, 1885.

[Frege(1914)] Gottlob Frege. Logic in mathematics. *Posthumous writings*, pages 203–250, 1914.

[Frege and Austin(1980)] Gottlob Frege and John Langshaw Austin. *The foundations of arithmetic: A logico-mathematical enquiry into the concept of number*. Northwestern University Press, 1980.

[Frege and MacGuinness(1984)] Gottlob Frege and Brian Francis MacGuinness. *Collected papers on mathematics, logic, and philosophy*. Blackwell, 1984.

[Frege et al.(2013)Frege, Ebert, and Rossberg] Gottlob Frege, Philip A Ebert, and Marcus Rossberg. *Gottlob Frege: Basic Laws of Arithmetic*, volume 1. Oxford University Press, 2013.

[Goldfarb(2001)] W Goldfarb. Frege's conception of logic. In: *Future pasts: The analytic tradition in twentieth century philosophy*, 2001.

[Heck(1993)] Richard G Heck. The development of arithmetic in frege's grundgesetze der arithmetik. *The Journal of Symbolic Logic*, 58(02):579–601, 1993.

[Horty(2007)] John Horty. *Frege on definitions: A case study of semantic content*. Oxford University Press, 2007.

[Jeshion(2001)] Robin Jeshion. Frege's notions of self-evidence. *Mind*, 110(440):937–976, 2001.

[May and Wehmeier(2016)] R May and K Wehmeier. The proof of hume's principle. *Ebert and Rossberg*, 2016.

[May(2012)] Robert May. What frege's theory of identity is not. *Thought: A Journal of Philosophy*, 1 (1):41–48, 2012.

[Shieh(2008)] Sanford Shieh. Frege on definitions. *Philosophy Compass*, 3(5):992–1012, 2008.

[Tappenden(1995)] Jamie Tappenden. Extending knowledge and fruitful concepts: Fregean themes in the foundations of mathematics. *Noûs*, 29(4):427–467, 1995.

[Weiner(1984)] Joan Weiner. The philosopher behind the last logicist. *The Philosophical Quarterly (1950-)*, 34(136):242–264, 1984.

[Weiner(1990)] Joan Weiner. *Frege in perspective*. Cornell University Press, 1990.

Strengthening Principles and Counterfactual Semantics

David Boylan and Ginger Schultheis*

Massachusetts Institute of Technology, Cambridge MA, USA
dboylan@mit.edu, vks@mit.edu

1 Introduction

There are two leading theories about the meaning of counterfactuals like (1):

(1) If David's alarm hadn't gone off this morning, he would have missed class.

Both say that (1) means, roughly, that David misses class in all of the closest worlds where his alarm doesn't go off. They disagree about how this set of worlds is determined. The *Variably Strict Analysis* (VSA) says that the domain *varies* from antecedent to antecedent. The *Strict Analysis* (SA) says it doesn't.[1] VSA and SA validate different inference patterns. For example, VSA validates *Antecedent Strengthening*, whereas SA does not. Early VSA theorists, such as Lewis (1973) and Stalnaker (1968), believed that certain apparent counterexamples to Antecedent Strengthening, which are now known as *Sobel Sequences*, refuted SA. More recently, defenders of SA have responded by enriching SA with certain *dynamic* principles governing how context evolves. They argue that Sobel sequences are not counterexamples to a *Dynamic Strict Analysis* (Dynamic SA).

But Antecedent Strengthening is just one of a family of strengthening principles. We focus on a weaker principle—*Strengthening with a Possibility*—and give a counterexample to it. The move to Dynamic SA is of no help when it comes to counterexamples to Strengthening with a Possibility. We show that these counterexamples are easily accommodated in a VSA framework, and we explain how to model our case and others like it using a Kratzerian ordering source.

2 Two Theories of Counterfactuals

Both VSA and SA assume that context supplies a comparative closeness ordering on worlds, represented by $\preceq_{c,w}$, which compares any two worlds with respect to their similarity to a world w.[2]

VSA uses a *selection function*, a contextually-determined function $f_{\preceq_c}$ that takes an antecedent A, and a world w, and returns the set of closest A-worlds to w, according to $\preceq_{c,w}$. A world w' is among the closest A-worlds to w just if there's no other A-world w'' that's closer to w than w' is. VSA's semantic entry for the counterfactual runs as follows:

VSA $\quad [\![A \boxright C]\!]^{c,w} = 1$ iff $\forall w' \in f_{\preceq_c}(A, w) : [\![C]\!]^{c,w'} = 1$.[3]

*The authors made equal contributions to the paper. Thanks to Justin Khoo, Matt Mandelkern, Milo Phillips-Brown, Bob Stalnaker, Kai von Fintel, and Steve Yablo for helpful discussion.

[1] For defenses of VSA, see Stalnaker (1968), Lewis (1973), Kratzer (1981a), Kratzer (1981b), Moss (2012), and Lewis (2017). For defenses of SA, see von Fintel (2001) and Gillies (2007).

[2] $\preceq_{c,w}$ is transitive, reflexive, antisymmetric, and at least weakly centered.

[3] This statement of VSA makes the limit assumption. (This is purely for ease of presentation.) It is neutral about the uniqueness assumption.

SA replaces the selection function with an accessibility relation $min_{\preceq_c}$ that takes a world w and returns the set of closest worlds to w, according to $\preceq_{c,w}$. (Unlike the selection function $f_{\preceq_c}$, $min_{\preceq_c}$ does not take the antecedent as argument.) Here's SA's semantic entry:

SA $\quad [\![A \boxright C]\!]^{c,w} = 1$ iff $\forall w' \in min_{\preceq_c}(w) \cap [\![A]\!]^c$: $[\![C]\!]^{c,w'} = 1$.

We said that VSA and SA do not validate all the same inference patterns. Here's one principle they disagree about:

(2) **Antecedent Strengthening**. $A \boxright C \vDash (A \wedge B) \boxright C$

SA validates Antecedent Strengthening, whereas VSA does not. It's not hard to see why. SA doesn't allow the evaluation domain to vary. If all the closest worlds where A is true are worlds where C is true, then *a fortiori*, all of the closest worlds where A *and* B are true are worlds where C is true. On the other hand, VSA allows the evaluation domain to vary from antecedent to antecedent: The closest AB-worlds need not be the closest A-worlds. So, what's true in all of the closest A-worlds may be false in some of the closest AB-worlds.

3 The State of Play

Antecedent Strengthening seems subject to counterexample. Consider this Sobel sequence:

(3) a. If I had struck the match, it would have lit.
b. But of course, if I had struck the match *and* it had been soaked in water last night, it wouldn't have lit.

Sentences (3-a) and (3-b) seem consistent. Indeed, in most ordinary match-striking scenarios, (3-a) and (3-b) are both true. But if Antecedent Strengthening is valid, (3) is not consistent. If it's true that if I'd struck the match, it would have lit, then, by Antecedent Strengthening, it follows that if I'd struck the match *and* it had been soaked in water, it would (still) have lit.

That it validates Antecedent Strengthening would seem to be a clear strike against SA. But things aren't quite so simple. As von Fintel (2001) shows, a suitably sophisticated strict conditional analysis *can* account for the sequence in (3). His strategy is to appeal to the dynamic effects of counterfactuals in conversation: Though (3-b) isn't true when (3-a) is uttered, *asserting* (3-b) changes the context so that it comes out true. More precisely: Counterfactuals presuppose that their domains contain some worlds where the antecedent is true. When that presupposition is not met, the context is minimally altered to ensure that it is. Suppose a speaker utters a counterfactual $A \boxright C$. If the domain contains A-worlds, nothing changes; if it doesn't, it *expands* to include the closest A-worlds. $A \boxright C$ is true in this *new* context just in case all of the A-worlds in the expanded set are C-worlds.

Let's apply von Fintel's Dynamic SA to our example. When the speaker asserts (3-a), the domain expands to include the closest worlds where she strikes the match. (3-a) is true. So in all of these worlds, the match lights. But any world where the match lights is one where the match is *dry*. So the presupposition of (3-b) isn't satisfied. When the speaker *utters* (3-b), the domain expands to include the closest worlds where she strikes the match and the match is wet. Since all of these worlds are ones where the match doesn't light, (3-b) comes out true.

Note that Antecedent Strengthening is not *classically* valid on Dynamic SA. An inference is classically valid just in case its conclusion is true whenever its premises are. Dynamic SA says that Antecedent Strengthening is merely *Strawson valid*: Whenever $A \boxright C$ is true, and

$(A \wedge B) \boxrightarrow C$ is *defined*, $(A \wedge B) \boxrightarrow C$ is true, too.[4] Dynamic SA allows contexts where $A \boxrightarrow C$ is true yet $(A \wedge B) \boxrightarrow C$ is undefined. This is critical to Dynamic SA's account of Sobel sequences. It is the fact that (3-b) is undefined, rather than simply false, that forces the context to change when (3-b) is asserted so that (3-b) comes out true.

We aim to advance the debate between VSA and Dynamic SA by looking at a broader range of data. Antecedent Strengthening is the strongest of a family of strengthening principles. By Strawson-validating Antecedent Strengthening, Dynamic SA predicts that a whole host of strengthening principles are Strawson-valid. We argue that this prediction is unwelcome. We focus on one strengthening principle—*Strengthening with a Possibility*—and present a counterexample to it. Dynamic SA *classically* validates this principle, rather than (merely) Strawson-validating it. This means that Dynamic SA's dynamic resources are of no help when it comes to counterexamples to Strengthening with a Possibility.

4 Strengthening with a Possibility

We can think of a strengthening principle as a principle that allows us to move from a counterfactual $A \boxrightarrow C$, along with certain auxiliary premises, to a counterfactual with a strengthened antecedent $(A \wedge B) \boxrightarrow C$. More formally, where $n \geq 0$, we have:

(4) **Strengthening Principle**. $A \boxrightarrow C, P_1, ..., P_n \vDash (A \wedge B) \boxrightarrow C$

Antecedent Strengthening is the instance of (4) where $n = 0$. It says that *no* further premises are needed to strengthen the antecedent of a counterfactual. This makes it the strongest strengthening principle: A semantics that validates it validates *every* strengthening principle. Classical validity is monotonic: *Adding* premises never turns a valid inference into an invalid one. Similar reasoning shows that *Strawson*-validating Antecedent Strengthening Strawson-validates every other strengthening principle—Strawson-entailment is monotonic.[5]

There are weaker strengthening principles that allow us to strengthen an antecedent not with just *any* conjunct, but only those that satisfy some auxiliary premises. We're interested in Strengthening with a Possibility:[6]

[4] The inference from $A, P_1, ..., P_n$ to C is Strawson-valid iff for any c such that $[\![A]\!]^{c,w_c}$, $[\![P_1]\!]^{c,w_c}$, ..., $[\![P_n]\!]^{c,w_c}$ and $[\![C]\!]^{c,w_c}$ are all defined and such that $[\![A]\!]^{c,w_c} = [\![P_1]\!]^{c,w_c} = ... = [\![P_n]\!]^{c,w_c} = 1$, $[\![C]\!]^{c,w_c} = 1$ also.

[5] **Proof**: Suppose that $A, P_1, ..., P_n \nvDash_{Str} C$. There there must be some c such that $[\![A]\!]^{c,w_c} = [\![P_1]\!]^{c,w_c} = ... = [\![P_n]\!]^{c,w_c} = 1$ but $[\![C]\!]^{c,w_c} = 0$. But then, since c itself is a context where $[\![A]\!]^{c,w_c} = 1$ but $[\![C]\!]^{c,w_c} = 0$, we have $A \nvDash_{Str} C$. Contraposing, if $A \vDash_{Str} C$ then $A, P_1, ...P_n \vDash_{Str} C$.

[6] Here we assume that $A \diamondrightarrow B$ is the dual of $A \boxrightarrow B$. So, according to Dynamic SA, it has the following semantics:

(i) $[\![A \diamondrightarrow B]\!]^{c,w} = 1$ iff $\exists w' \in min_{\preceq_c}(w) : [\![A]\!]^{c,w'} = 1$ and $[\![B]\!]^{c,w'} = 1$.

And according to VSA it has the following semantics:

(ii) $[\![A \diamondrightarrow B]\!]^{c,w} = 1$ iff $\exists w' \in f_{\preceq_c}(A, w) : [\![B]\!]^{c,w'} = 1$.

Throughout we also assume that English *might*-counterfactuals have the semantics of $A \diamondrightarrow B$. This assumption is called *Duality*. Gillies and von Fintel seem to accept Duality. Note also that Duality falls out of the widely-accepted restrictor analysis of conditionals in Kratzer (1986): on this analysis, the 'might' will only quantify over worlds that make the antecedent true and so *might*-counterfactuals will have the truth-conditions of $\diamondrightarrow$.

That being said, Duality has been denied by some in the wider literature on counterfactuals (in particular, by various defenders of Counterfactual Excluded Middle like Stalnaker (1981) and Williams (2010)). We assume Duality merely for ease of exposition. Our central counterexample can be stated without it. See footnote 8 for further details.

(5) **Strengthening with a Possibility**. $(A \boxrightarrow C) \wedge (A \diamond\!\!\rightarrow B) \vDash (A \wedge B) \boxrightarrow C$

(7) says that one can strengthen an antecedent with any proposition with which that antecedent is *counterfactually consistent*. Suppose it's true that if I'd taken modal logic next semester, I would have passed. Does that mean that I would have passed had I taken the class and the class was taught by Joe? According to Strengthening with a Possibility, that depends on whether Joe *might* have been the teacher, had I taken the class. If Joe couldn't have taught the class—say, because he was on leave—I can truly say that I would have passed even if I would have bombed a class taught by Joe. On the other hand, if Joe might have taught the class, then I can't truly say that I would have passed unless I would have passed Joe's class, too.

We said that Antecedent Strengthening is the strongest strengthening principle. So, by Strawson-validating Antecedent Strengthening, Dynamic SA Strawson-validates Strengthening with a Possibility. But we can show something stronger: By Strawson-validating Antecedent Strengthening, Dynamic SA *classically* validates Strengthening with a Possibility.[7] This is important. If Strengthening with a Possibility is classically valid, we can't appeal to the dynamic resources of Dynamic SA to account for apparent counterexamples. By the definition of classical validity, $(A \wedge B) \boxrightarrow C$ is *defined* (and true) in any context in which $A \boxrightarrow C$ and $A \diamond\!\!\rightarrow B$ are true. But if $(A \wedge B) \boxrightarrow C$ is defined, then *asserting* $(A \wedge B) \boxrightarrow C$ won't change the context. The domain will not expand to make $(A \wedge B) \boxrightarrow C$ false, as we would hope; $(A \wedge B) \boxrightarrow C$ will simply come out true in the original context in which $A \boxrightarrow C$ and $A \diamond\!\!\rightarrow B$ are uttered.

5 Against Dynamic SA

In this section, we present an apparent counterexample to Strengthening with a Possibility. Here it is:

> *Dice*: Alice, Billy, and Carol are playing a simple game of dice. Anyone who gets an odd number wins $10; anyone who gets even loses $10. Each player throws their dice. Alice gets odd; Billy gets even; and Carol gets odd.

Now consider this sequence of counterfactuals:

(6) a. If Alice and Billy had thrown the same type of number, then at least one person would still have won $10.
b. If Alice and Billy had thrown the same type of number, then Alice, Billy and Carol could have *all* thrown the same type of number. (So they could have all won $10.)
c. If Alice, Billy and Carol had all thrown the same type of number, then at least one person would still have won $10.

(6-a) and (6-b) seem true, but (6-c) is dubious. (6-a) seems right because if Alice and Billy had thrown the same type of number, nothing would have changed with respect to *Carol*—she'd still have rolled odd. So someone would still have won $10.

(6-b) seems right, too. If Alice and Billy had thrown the same type of number, either Alice or Billy would have gotten a different number from the one they actually got. But there's no reason to think it would have been Alice rather than Billy: Billy might have thrown odd, along with Alice and Carol.

[7] The proof is straightforward. Suppose that for a given context c, (i) $A \boxrightarrow C$ is true in c, and (ii) $A \diamond\!\!\rightarrow B$ is true in c. It follows from (ii) and Dynamic SA that the domain in c contains worlds where A and B are both true. But that's just to say (iii) that $(A \wedge B) \boxrightarrow C$ is *defined* in c. Since Antecedent Strengthening is Strawson-valid, (i) and (iii) entail that $(A \wedge B) \boxrightarrow C$ is true in c.

But (6-c) seems wrong. There are two ways for Alice, Billy, and Carol to throw the same type of number. They could all roll odd or they could all roll even. And we can't just rule out the latter. If Alice, Billy, and Carol had thrown the same type of number, they might have all thrown even, so there might have been no winner: (6-c) is false.

Dynamic SA wrongly predicts that (6-c) follows from (6-a) and (6-b). For (6-a), (6-b), and (6-c) are respectively equivalent to:[8]

(7) a. *Alice Billy same* $\Box\!\!\rightarrow$ *someone wins \$10*
b. *Alice Billy same* $\Diamond\!\!\rightarrow$ (*Alice Billy same* $\wedge$ *Billy Carol same*)
c. (*Alice Billy same* $\wedge$ *Billy Carol same*) $\Box\!\!\rightarrow$ *someone wins \$10*

Suppose (7-a) and (7-b) are true. Since (7-b) is true, some worlds in the domain are ones where its antecedent and consequent are true—that is, where Alice, Billy, and Carol all throw the same type of number. But that's just to say that (7-c) is *defined*. Dynamic SA Strawson-validates Antecedent Strengthening. So, if (7-a) is true, and (7-c) is defined, then (7-c) must be true, too. That's wrong. (7-a) and (7-b) are true, and (7-c) is not.

6 A Way Out?

In its current form, Dynamic SA cannot account for our judgments about these sentences. Is there a way to modify Dynamic SA so that it can? We don't think so. Let us explain.

We know that someone wins just in case someone rolls odd. To predict that (6-a) is true, there must be someone who rolls odd in all domain-worlds where Alice and Billy roll the same type of number. And to predict that (6-c) is false, some domain-worlds where Alice and Billy (and Carol) roll the same type of number must be ones where everyone rolls *even*. So, the domain must expand between utterances of (6-a) and (6-c). It must start out containing no worlds where Alice, Billy, and Carol roll even, but acquire some by the time we get to (6-c). There are only two ways for this to happen. Either asserting (6-b) expands the domain, or asserting (6-c) does. We already ruled out the latter—if (6-a) and (6-b) are true, (6-c) is true, and thereby defined, so asserting (6-c) will not expand the domain. So if *anything* expands the domain, it must be asserting (6-b).

(6-b) is a *might*-counterfactual. We haven't yet said how they update the domain. One possibility is that they work just like *would*-counterfactuals do: (6-b) presupposes that the domain contains worlds Alice and Billy roll the same type of number. But this account won't help with our data. (6-a) and (6-b) have the same antecedent, so there can be no shifting that is triggered by the latter that isn't already triggered by the former. A different idea can be found in Gillies (2007). Gillies argues that $A \Diamond\!\!\rightarrow B$ presupposes that the domain contains worlds where A and B are both true.[9] For example, (6-b) presupposes that the domain contains

[8] In assuming that (6-b) is equivalent to (7-b), we assume Duality. However, as we noted, the counterexample does not ultimately rely upon it. We can state the dual of the *would*-counterfactual using wide-scope negation:

(i) a. If Alice and Billy had thrown the same type of number, then at least one person would still have won \$10.
b. It's not true that if Alice and Billy had thrown the same type of number, then Alice, Billy and Carol *wouldn't* have all thrown the same type of number.
c. If Alice, Billy and Carol had all thrown the same type of number, then at least one person would still won \$10.

We notice no difference in our judgements here.

[9] Gilies does not think this assertability condition is a genuine presupposition, even though he calls it an 'entertainability presupposition'. We take no view on how to cash out entertainability presuppositions.

worlds where Alice, Billy, and Carol all roll the same.

How might Gillies' theory help with *Dice*? Suppose that the initial context is such that, in all domain-worlds, Alice and Carol roll odd, and Billy rolls even. (6-a)'s presupposition isn't met in this context, so asserting (6-a) expands the domain, adding worlds where Alice and Billy roll the same. Suppose we include worlds where Alice and Billy roll even, but none where they roll odd. (We can't include any worlds where *Carol* rolls even, lest we render (6-c) false.) But in that case, (6-b)'s presupposition won't be met. (6-b)'s consequent is true only if Alice, Billy, and Carol roll the same. But, as we've set things up, the domain doesn't contain any worlds where they all roll the same. This means that asserting (6-b) will add worlds where Alice, Billy, and Carol all throw the same type of number. If we include worlds where they all throw even, (6-c) comes out false.

So far things are looking better for Dynamic SA. But trouble is near. If (6-b) introduces worlds where Alice, Billy, and Carol all throw even, we will indeed make (6-c) false, but there are other, less welcome consequences. Consider the sequence:

(6-b) If Alice and Billy had rolled the same type of number, Alice, Billy, and Carol might have *all* rolled the same type of number.

(8) If Alice and Billy had rolled the same type of number, Carol might not have rolled odd.

(6-b) is true, but (8) is false. Indeed, (8) is false for the same reason that (6-a) is true—there's no reason to suppose that, if Alice and Billy had rolled the same type of number, things might have changed with respect to *Carol*. She would have still rolled odd. But if (6-b) adds to the domain worlds where Alice, Billy, and Carol throw even, (8) will come out true.[10]

We don't want (6-b) to add worlds where Alice, Billy, and Carol all roll even. When we evaluate (6-b), we're still holding fixed that Carol rolls odd—that's why we judge (8) false. (We judge (6-b) true not because we think they might have all thrown even, but because we think they might have all thrown odd.) To be sure, things change by the time we get to (6-c). At that point, we are considering worlds where they all throw even—we judge (6-c) false because they might have all thrown even and lost. But it isn't (6-b) that makes those worlds relevant. It is only when we hear (6-c) that we consider worlds where Carol rolls even.

We've now seen that asserting (6-b) doesn't expand the domain, and neither does asserting (6-c). But if there's no domain expansion between (6-a) and (6-c), Dynamic SA cannot predict a false reading of (6-c).

7 Variably Strict Semantics

By its very structure, SA is committed to Strengthening with a Possibility. No assumptions about its underlying closeness relation were needed to prove this. Not so for VSA. Strengthening with a Possibility is not written into the semantics of VSA; rather, it corresponds to a certain formal constraint on the closeness ordering $\preceq_{c,w}$, *almost-connectedness*. Some of VSA's proponents, including Stalnaker (1968) and Lewis (1973), do enforce this constraint.[11] We show

[10] We can make this same point with the following *would*-counterfactual:

(i) If Alice and Billy had rolled the same type of number, Carol would still have rolled odd.

(i) is intuitively true. But if (6-b) introduces worlds where Alice, Billy, and Carol roll even, (i) will be false.

[11] In particular, both Stalnaker (1968) and Lewis (1973) say that whatever else is true about the ordering on worlds, it is total. Total orderings rule out incomparabilities of any kind, and so do not allow for failures of almost-connectedness.

that by adding a Kratzerian *ordering source* to the semantics we naturally generate an ordering without this constraint, allowing us to predict the counterexamples in a principled way.[12]

7.1 Predicting the counterexamples

Say that the closeness ordering $\preceq_w$ is *almost-connected* just in case $\forall w_1, w_2, w_3 : (w_1 \prec_w w_2 \rightarrow (w_1 \prec_w w_3) \vee (w_3 \prec_w w_2))$. If w_1 is closer to w than w_2 is, then for any third world w_3, either w_1 is closer to w than w_3 is, or w_3 is closer to w than w_2 is. Simplifying, if w_1 beats w_2, then either w_1 beats w_3, or w_3 beats w_2. Where $\preceq_w$ is a partial order, Strengthening with a Possibility is valid just in case $\preceq_w$ is almost-connected.[13]

To predict the data in *Dice*, it's not enough that the ordering simply fail to be almost-connected; it must fail to be almost-connected *in the right ways*. To predict (6-a), we need worlds where Alice and Billy get even and Carol gets odd to be closer to the actual world than worlds where they all get even. To predict (6-b), we need worlds where all three get odd to be among the closest worlds to actuality where Alice and Billy get the same of type of number. And, finally, to predict (6-c), we need worlds where they all get odd to *not* be closer to actuality than worlds where they all get even. These three jointly hold just in case worlds where they all get odd are neither closer to actuality than worlds where they all get even, nor further away from actuality than worlds where Carol gets odd but Alice and Billy get even.

How do we guarantee that context supplies an ordering with this structure? Our suggestion is to follow Kratzer (1981a) and Kratzer (1981b) and posit an extra contextual parameter—an *ordering source*, a function that takes a world w and returns a set of propositions. This set of propositions represents the facts about w that the speakers judge relevant to determining similarity. We then define our ordering in terms of those propositions. w_1 is at least as close to w as w_2 is just if it makes true all the same ordering source propositions as w_2, and possibly more. Formally:

(9) $w_1 \preceq_w w_2$ iff $\{p \in g(w) : w_1 \in p\} \supseteq \{p \in g(w) : w_2 \in p\}$

w_1 is at least as close to w as w_2 is just in case every proposition in $g(w)$ that is true in w_2 is also true in w_1. w_1 is strictly closer to w than w_2 is just in case, every proposition in $g(w)$

[12] Kratzer also adds to her semantics a modal base which is shifted by the antecedent. We omit this in what follows for ease of exposition, and make the simplifying assumption that it is true in all worlds that one wins the game if and only if one rolls odd.

We also depart from Kratzer with respect to which facts we think the ordering source holds fixed. For Kratzer, the ordering source is totally realistic: the intersection of $g(w)$ is just $\{w\}$. We do not make this assumption; instead, we only include the facts that are relevant, in the sense spelled out in 7.2. Were we to spell out the semantics in full, we would say that all other relevant details about the case, such as the rules of the game, go in the modal base, rather than in the ordering source.

[13] **Proof**: $\Rightarrow$: Our model that follows demonstrates that if Strengthening with a Possibility is valid, then $\preceq$ is almost-connected. If a frame is not almost-connected, then we can build a model on it like the one in the text.

$\Leftarrow$: Suppose that $\preceq$ is almost-connected and suppose that, for contradiction, that Strengthening with a Possibility is not valid. Then there is some world w_1 such that $A \boxrightarrow C$ and $A \diamond\!\!\rightarrow B$ are true there but $A \wedge B \boxrightarrow C$ is not. This means that $f(A, w_1) \subseteq C$, there is a world $w_2 \in f(A, w_1)$ such that B is true at w_2 and there is a world $w_3 \in f(A \wedge B, w)$ such that $\neg C$ is true there. w_3 cannot be in $f(A, w)$: unlike w_3 all worlds in $f(A, w)$ are C worlds. By definition of f, this means that there must be some world w_4 in $f(A, w)$ such that $w_4 \prec_{w_1} w_3$.

Now consider whether either $w_4 \prec_{w_1} w_2$ or $w_2 \prec_{w_1} w_3$. In fact, the first disjunct cannot hold: by definition of f, if it did then w_2 would not be in $f(A, w_1)$ after all. But the second disjunct cannot be true either. Again by definition of f, if it were then w_3 would not be in $f(A \wedge B, w_1)$. But now we have proved that, contrary to our supposition that $\preceq$ is not almost connected: $w_4 \prec_{w_1} w_3$ but neither $w_4 \prec_{w_1} w_2$ nor $w_2 \prec_{w_1} w_3$. So if $\preceq$ is almost-connected Strengthening with a Possibility must be valid. (To the best of our knowledge, this result was first shown by Veltman (1985).)

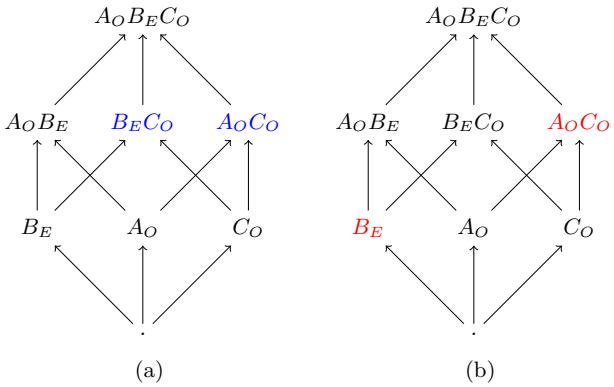


Figure 1

that's true in w_2 is true in w_1, and some proposition in $g(w)$ that's true in w_1 is false in w_2.

In our example, the relevant facts are those that concern who got what type of number. We assume, then, that the ordering source is that in (10):

(10) $g(w) = \{\textit{Alice gets odd}, \textit{Billy gets even}, \textit{Carol gets odd}\}$

Let A_O, B_E, and C_O be the propositions that Alice rolls odd, Billy rolls even, and Carol rolls odd, respectively. The ordering source in (10) gives rise to the ordering in Figure 1. In the top-ranked worlds, things are just as they actually are—Alice and Carol roll odd, and Billy rolls even. Next we have worlds where things differ in *one* respect—worlds where either Alice or Carol rolls even instead of odd, or Billy rolls odd instead of even. Then we have worlds differing in *two* respects, and finally, worlds where *everything* is different—Alice and Carol roll even, and Billy rolls odd.

Let's see how VSA predicts the right judgments in *Dice* using this ordering. The closest worlds where Alice and Billy throw the same type of number are in blue in Figure 1(a). In both worlds, Carol rolls odd and wins \$10, so (6-a) is true: If Alice and Billy had rolled the same, one person would still have won \$10. Moreover, in one of the closest worlds where Alice and Billy roll the same, Alice, Carol, and Billy all roll odd. So (6-b) is also true: If Alice and Billy had rolled the same, Alice, Billy, and Carol might have all rolled the same.

Finally, turn to (6-c), which says that if Alice, Billy, and Carol had rolled the same, someone would still have won \$10. We find the worlds where Alice, Billy, and Carol roll the same type of number. They are highlighted in red in Figure 1(b). We have worlds where Alice, Billy, and Carol all throw odd (top right) and worlds where they all throw even (bottom left). These are incomparable—neither is closer to actuality than the other is. (The reason they are incomparable is that the sets of ordering source propositions true at each are disjoint.) Both are among the closest worlds where Alice, Billy, and Carol roll the same. So (6-c) is false: In some of the closest worlds where they throw the same, they throw even, and nobody wins.

7.2 Other Cases

We've argued that Strengthening with a Possibility has counterexamples, and we've offered a Kratzerian premise semantics that doesn't validate it. Still, the inference often *seems* valid. Suppose I say, with confidence, that if I had taken modal logic last semester, I would have

passed. You reply that if I had taken the course, it might have been taught by Joe, who's notorious for his difficult problem sets and harsh grading. If I accept your response, I seem to have two options. I could stand firm, insisting that I would have passed even Joe's challenging course, or I could retreat, rescinding my earlier claim that I would have passed the class. What I *can't* do is maintain that I would have passed the course, even though I might not have passed a course taught by Joe. That I don't have this option is only explained if Strengthening with a Possibility does not fail in this particular case. We must place certain constraints on when the inference can fail. We want it to fail in *Dice*, but not here.

Our idea is to place a constraint on how our ordering sources relate to salient questions in context. Say that Q_c is the most refined salient (non-counterfactual) question in c. Now let $g^{\neg}$ be the following function: $g^{\neg}(w) = \{p : \neg p \in g(w)\}$; that is, $g^{\neg}(w)$ contains all and only the negations of propositions in $g(w)$. Finally consider all the sets of maximal consistent propositions built out of $g(w) \cup g^{\neg}(w)$; call it G_w. We propose the following constraint on ordering sources: whatever g_c is, it must be the case that $G_{w_c} = Q_c$. This constraint tells us that the ordering source cannot distinguish between worlds in ways that are not already present in the most refined salient question.

With this constraint in hand, we can prove that we get failures of Strengthening with a Possibility only if there are two distinct answers to Q_c that realise the antecedent of the final strengthened conditional.

> There are A, B, C such that $[\![A \boxright C]\!]^{c,w,g} = 1$, $[\![A \diamond\!\!\rightarrow B]\!]^{c,w,g} = 1$ and $[\![A \wedge B \boxright C]\!]^{c,w,g} = 0$ only if $\exists p, q \in Q_c : p \vDash A \wedge B$ and $q \vDash A \wedge B$ and $p \neq q$.[14]

Put informally, the reason for this is as follows: if there is only one partition cell, call it p, that entails $A \wedge B$, then either p is a subset of the closest A-worlds or not. If it is, then, if $A \boxright C$ is true, all the closest $A \wedge B$-worlds will have to be C-worlds. If it isn't, then, since all worlds in p are equally close, no B-worlds will be among the closest A-worlds and so $A \diamond\!\!\rightarrow B$ will be false.

To see how this helps, let us return to the case of Joe. Here, quite plausibly the most salient question is *Did I take logic? And did Joe teach?*, which gives us the following partition:

> {*I take logic and Joe teaches*, *I take logic and Joe doesn't teach*,
> *I don't take logic and Joe teaches*, *I don't take logic and Joe doesn't teach*}

There is only one cell of the partition which makes true our strengthened antecedent, namely, *I take logic and Joe teaches*. Given our result from above, we can see that the relevant instance of Strengthening with a Possibility will go through.

Here we see yet another advantage of our premise semantics. Not only does it offer an account of when Strengthening with a Possibility fails, it also offers an explanation of why it

[14] **Proof.** Suppose that $A \boxright C$, $A \diamond\!\!\rightarrow B$, but $A \wedge B \boxright C$ is false. For contradiction, suppose there's just one cell that makes $A \wedge B$ true. Call it Q. We appeal to three facts:

1. All worlds in a partition cell are equally good. This is because they all make the same ordering source propositions true.
2. $Q = Q \cap A = Q \cap (A \wedge B)$ This is because Q already contains only $A \wedge B$ worlds.
3. $Q \cap (A \wedge B) = f(A \wedge B, w)$ This follows from the definition of f plus the fact that Q is the only $A \wedge B$ cell.

Either $Q \cap A \subseteq f(A, w)$ or it isn't. Suppose it is. Then, by facts 2 and 3, $f(A \wedge B, w) \subseteq f(A, w)$. And since $f(A, w) \subseteq C$, $A \wedge B \boxright C$ is true, contrary to our supposition. Suppose it isn't. Then $f(A, w)$ contains *no* B-worlds: Q contains the only $A \wedge B$ worlds and, by fact 1, they are all equally good. So $A \diamond\!\!\rightarrow B$ is false, contrary to our supposition.

often seems to go through. In cases like ours where Strengthening with a Possibility fails, we are interested in different ways in which the antecedent could be true. But in normal, simple cases, we do not make such fine distinctions and so the inference seems valid.

8 Conclusion

We suggested that the debate between SA and VSA could be clarified by looking at a wider range of strengthening principles. This suggestion has been borne out. Dynamic SA validates Strengthening with a Possibility. But this inference is not valid. Counterexamples to Strengthening with a Possibility pose a much more serious problem for Dynamic SA than counterexamples to Antecedent Strengthening itself. While Antecedent Strengthening is merely Strawson-valid, Strengthening with a Possibility is *classically* valid. Counterexamples to it do not involve presupposition failure, so the dynamic principles that drive context change do not apply. But if that's right, Dynamic SA has no way to account for counterexamples to Strengthening with a Possibility. VSA, on the other hand, can easily model failures of Strengthening with a Possibility. We conclude that the failure of Strengthening with a Possibility tells strongly against Dynamic SA and in favor of an ordering source-based version of VSA.

References

Anthony Gillies. Counterfactual scorekeeping. *Linguistics and Philosophy*, 30(3):329–360, 2007.

Angelika Kratzer. The notional category of modality. *Words, worlds, and contexts*, pages 38–74, 1981a.

Angelika Kratzer. Partition and revision: The semantics of counterfactuals. *Journal of Philosophical Logic*, 10(2):201–216, 1981b.

Angelika Kratzer. Conditionals. *Chicago Linguistics Society*, 22(2):1–15, 1986.

David Lewis. *Counterfactuals*. Blackwell, 1973.

Karen Lewis. Counterfactual discourse in context. *Noûs*, 50(4), 2017.

Sarah Moss. On the pragmatics of counterfactuals. *Noûs*, 46(3):561–586, 2012.

Robert Stalnaker. A theory of conditionals. *Americal Philosophical Quarterly*, pages 98–112, 1968.

Robert Stalnaker. A defense of conditional excluded middle. In William Harper, Robert C. Stalnaker, and Glenn Pearce, editors, *Ifs*, pages 87–104. Reidel, 1981.

Frank Veltman. *Logics for Conditionals*. PhD thesis, University of Amsterdam, 1985.

Jonathan Vogel. Are there counterexamples to the closure principle? In Michael David Roth and Glenn Ross, editors, *Doubting: Contemporary Perspectives on Skepticism*, pages 13–29. Dordrecht: Kluwer Academic Publishers, 1990.

Kai von Fintel. Counterfactuals in a dynamic context. *Current Studies in Linguistics Series*, 36:123–152, 2001.

J. Robert G. Williams. Defending conditional excluded middle. *Noûs*, 44(4):650–668, 2010.

Expressing Agent Indifference in German*

Brian Buccola[1] and Andreas Haida[2]

[1] Laboratoire de Sciences Cognitives et Psycholinguistique (ENS, EHESS, CNRS), Département d'Études Cognitives, École Normale Supérieure, PSL Research University, Paris, France
brian.buccola@gmail.com

[2] The Edmond and Lily Safra Center for Brain Sciences, The Hebrew University of Jerusalem, Jerusalem, Israel
andreas.haida@gmail.com

Abstract

The German indefinite modifier *irgend-* can give rise to agent indifference (AI) readings. We propose a novel account of AI that builds on the observation that the adverbial *einfach* 'simply' emphasizes the AI reading of *irgend-*. We assume that *einfach* references a simplicity order that determines, in relative terms, what is simple for the agentive subject of the host sentence. For *irgend-*, we employ the by now standard assumption that it comes with a covert domain variable and activates subdomain alternatives. To derive AI, we argue that, if an agent has options for an action and preferences about which option to realize, then realizing one of many options (e.g. buying a single book from a large domain) is more complex than realizing one of fewer options (e.g. buying a single book from a subdomain). To create a link between the simplicity order referenced by *einfach* and the preference order employed in the derivation of AI, we show that the subdomain alternatives activated by *irgend-* can be associated with decision problems, and that these decision problems are equally simple iff the decision maker doesn't have preferences as to which of the expressed options to realize. We also compare German *irgend-* to Spanish *cualquiera* and to English *any* and discuss the consequences of our analysis for the theory of polarity sensitivity.

1 Introduction

If someone bought a book and did so randomly, or without any preference as to the choice of book, then we can say that the agent of this event was indifferent about the type and specimen of book she bought. Perhaps surprisingly, there are languages in which such *agent indifference* (AI) can be expressed without mentioning randomness of action or preferences about outcomes. For example, German and Spanish can express AI by means of certain indefinites that signify what the indifference is about, i.e. here the object of the book buying action. This is illustrated in (1) for German and (2) for Spanish (see [1, 2], henceforth AOMB).

(1) *Hans hat irgend-ein Buch gekauft.*
Hans has IRGEND-a book bought
'Hans bought a random book.'

(2) *Juan compró un libro cualquiera.*
Juan bought a book CUALQUIERA
'Juan bought a random book.'

As indicated by the glosses, the object expressions of (1) and (2) would be ordinary indefinites were it not for the modifiers *irgend-* and *cualquiera*, respectively. By means of these modifiers,

*We thank Luis Alonso-Ovalle, Luka Crnič, and Yosef Grodzinsky for discussing with us our ideas about *irgend-* and *einfach*, as well as two anonymous reviewers for the Amsterdam Colloquium for their comments. This research received funding from the Israel Science Foundation (ISF 1926/14), from the German-Israeli Foundation for Scientific Research (GIF 2353), and from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement n. 313610, and was supported by ANR-10-IDEX-0001-02 PSL* and ANR-10-LABX-0087 IEC.

(1) and (2) can convey that the agent bought a book (*actuality inference*), and that, as far as the agent's preferences go, it could've been any book (*indifference inference*). The preceding paraphrase, which is due to [1], characterizes AI in terms of a modal statement that references the preferences of the agent.[1,2] AOMB argue for Spanish that the modal aspect of AI is best analyzed as being hardwired into the meaning of *cualquiera*. In this paper, we propose an analysis for German *irgend-* that is crucially different: *irgend-* doesn't have a modal meaning component; rather, the modal inference arises through the interplay of the standard existential meaning of *irgendein Buch*, the alternatives that it activates, and an overt or covert modal operator that acts on those alternatives.[3]

Our approach is motivated by the observation that the AI reading of (1) can be emphasized with the help of the adverbial *einfach* 'simply': while (1) can be used by a speaker to convey not AI, but rather that she doesn't know or care to tell which book Hans bought (see e.g. [13]), (3) cannot be (easily) used to this effect.[4]

(3) *Hans hat einfach irgend-ein Buch gekauft.*
Hans has simply IRGEND-a book bought
'Hans simply bought a random book.'

(4) *Hans hat* Lolita *einfach raubkopiert.*
Hans has *Lolita* simply pirated
'Hans simply pirated *Lolita*.'

We will explicate the meaning contribution of *einfach* on the basis of an elementary case, viz. the sentence in (4). This sentence licenses a *simplicity inference*: given that buying and borrowing are alternatives of pirating, we can infer from (4) that buying or borrowing *Lolita* would not have been simpler for Hans. Simplicity is a context-sensitive relation over alternatives; here, pirating is simpler for Hans than buying or borrowing because, e.g., it requires less effort. What we claim then is that the modal inference of (1)/(3) is also a simplicity inference, and furthermore, that this simplicity inference entails AI.

We begin by detailing the denotation of *einfach* (§2). We then derive the truth conditions of (1)/(3) on the basis of our semantics for *einfach* and well-established assumptions about *irgend-* (§3). We proceed by making explicit the assumptions on which simplicity inferences entail AI. In a nutshell, we argue that an agent's preferences regarding (the outcome of) an action determine what is simple for the agent. The simplicity inference of sentences like (1)/(3) is only compatible with a preferenceless (i.e. indifferent) agent. We give a decision-theoretic account for the link between having (no) preferences and simplicity (§4). We then present further empirical support for our analysis. We show that the analysis of German *irgend-* must be crucially different from the analysis of Spanish *cualquiera* and that we must reject the assumption that *irgend-* is a negative polarity item (pace [3]). Furthermore, we show that in NPI licensing environments English *any* can convey AI, calling into question [3]'s assumptions about NPI licensing (§5).

2 The Denotation of *einfach*

We ultimately want to analyze the AI reading of (1)/(3) as resulting from an interaction between (a possibly covert version of) the adverbial *einfach* and the indefinite *irgendein*. Thus, we first

[1] [2] assumes that the modal relation is determined by the agent's goals, where the goals of a volitional agent bear a relation to decisions to act. Our analysis inherits from [2] the idea that there's a link between an agent's attitudes, for us her preferences, and her decision making.

[2] Randomness of action/agent indifference has also been analyzed in terms of counterfactuality. See [9, 5] for specific proposals and AOMB for discussion.

[3] See [1, 3] for analyses of this type and §5 for how and why our analysis deviates from its predecessors.

[4] We won't discuss whether (3) can be used at all to convey speaker ignorance or indifference instead of AI. To be clear, (3) is certainly *compatible* with these speaker attitudes, but we doubt that it can *convey* them.

need to explicate the semantic import of *einfach* on the basis of a simpler sentence like (4). To that end, the meaning of (4), we claim, consists of the following four parts:

(5) a. Hans pirated *Lolita*. (*actuality inference*)
b. Hans didn't buy or borrow *Lolita*. (*exhaustivity inference*)
c. Hans could've bought or borrowed *Lolita*. (*circumstantial possibility inference*)
d. Buying or borrowing *Lolita* wouldn't have been simpler for H. (*simplicity inference*)

We take it that *einfach* operates on alternatives, in a sense to be made precise shortly, and that the identification of the relevant alternatives is context-sensitive. In (5) we assume for concreteness that the relevant alternatives to pirating *Lolita* are buying it and borrowing it.

Evidence for these inferences comes in the form of falsity or infelicity judgments regarding (4) in contexts that don't support the relevant inference. For example, if Hans didn't acquire *Lolita* at all, then (4) is clearly false. Similarly, suppose that Hans is rich and lives next door to a bookstore (hence, buying *Lolita* would be very easy for him), and that Hans is also computer illiterate (hence, pirating *Lolita* would be very difficult for him). Suppose furthermore that Hans nevertheless decided to exert great effort in learning how to pirate *Lolita*. Then (4) is judged to be very odd.

In addition, we observe that the simplicity inference, (5d), is always relative to the agentive subject, here Hans, hence 'simpler *for Hans*'. To see why, consider a context where Hans has no money to buy *Lolita*, no transportation to the bookstore, etc. (and no computer to pirate it), but his friend Marie owns a copy of *Lolita*. Suppose further that Marie needs to read *Lolita* for class, but that she could easily lend Hans money to buy his own copy. In this case, to acquire *Lolita*, it's simpler *for Hans* that he borrow it from Marie, but simpler *for Marie* that she lend him money to buy it. If the simplicity inference associated with *einfach* could be relative to any contextually salient person, then (6) and (7) below could each be judged felicitous and true (or felicitous but false) depending on which person (Hans or Marie) the simplicity relation were relative to (and depending on whether Hans in fact bought or borrowed *Lolita*). However, only (7) is felicitous (and true if Hans did in fact borrow it; false otherwise); (6) is infelicitous (even if Hans did buy it).

(6) *Hans hat* Lolita *einfach gekauft.*
Hans has *Lolita* simply bought
'Hans simply bought *Lolita*.'

(7) *Hans hat* Lolita *einfach ausgeliehen.*
Hans has *Lolita* simply borrowed
'Hans simply borrowed *Lolita*.'

With these remarks in mind, we turn to the meaning contribution of *einfach*. For ease of exposition, we assume that *einfach* is a sentence adverb. Then the LF structure in (8) provides a suitable basis for our analysis of (4): *einfach* is coindexed with the subject (to capture the dependence of *einfach* on the agentive subject), the main verb induces alternatives by being focus-marked, and *einfach* has an exhaustification operator, *exh*, in its immediate scope (see, e.g., [10, 4]).[5]

(8) einfach$_i$ [$_{S_2}$ exh [$_{S_1}$ Hans$_i$ *Lolita* raubkopiert$_F$ hat]]

We propose that *einfach* denotes a modal operator that is restricted by a circumstantial modal base and an ordering source that characterizes what is simple for the (agentive) subject of the host sentence. The denotation of *einfach*$_i$ S is given in (9), where a is a variable assignment function, f the modal base (conceived of as a function from D_s to D_{st}), and g the ordering source (conceived of as a function from $D_e \times D_s$ to $D_{(st)t}$). The relation $>_{g(x,w)}$ ('simpler for x in w') is defined in (11) on the basis of the non-strict ordering relation in (10) (cf. [16]).

[5] For the main argument, nothing hinges on this implementation of exhaustification. Alternatively, one could assume that exhaustification is part of the meaning of *einfach* itself.

(9) $[\![\text{einfach}_i\ S]\!]^{a,f,g}(w) = 1$ iff $[\![S]\!]^a(w) = 1$ and $\neg\exists p \in \text{Alt}(S) : p \cap f(w) >_{g(a(i),w)} [\![S]\!]^a \cap f(w)$

(10) $\forall p, q \in D_{st} : p \geq_{g(x,w)} q$ iff for every p-world u, there is a q-world v such that $u \geq_{g(x,w)} v$
$\forall u, v \in D_s : u \geq_{g(x,w)} v$ iff $\{p \in g(x,w) : p(v) = 1\} \subseteq \{p \in g(x,w) : p(u) = 1\}$

(11) $\forall p, q \in D_{st} : p >_{g(x,w)} q$ iff $p \geq_{g(x,w)} q$ and $q \not\geq_{g(x,w)} p$

We assume that in (8) *exh* and *einfach* operate on the alternatives in (12) and (13), respectively. Then the actuality inference, (5a), and the exhaustivity inference, (5b), both follow since *einfach* S_2 asserts that S_2 is true, where S_2 denotes the exhaustified meaning of S_1 relative to the alternatives in (12), viz. that Hans pirated but didn't buy or borrow *Lolita*. The circumstantial possibly inference, (5c), follows from the 'no alternative is simpler' condition in (9) since $p \cap f(w) >_{g(x,w)} q \cap f(w)$ is trivially true (for all g, x, w, q) if $p \cap f(w) = \varnothing$ (i.e. if an alternative p is not circumstantially possible). The simplicity inference, (5d), follows from the 'no alternative is simpler' condition in conjunction with (13). Importantly, if $\text{Alt}(S_2)$ were not a set of exhaustified alternatives, then this condition would be trivially satisfied: for instance, the proposition $[\lambda w\,.\,\text{Hans bought } \textit{Lolita} \text{ in } w] \cap f(w)$ contains worlds in which Hans bought, pirated, and borrowed *Lolita* (given circumstances that don't rule out the possibility of acquiring *Lolita* in several ways), and such a world cannot be any simpler than any world in which Hans pirated *Lolita* (even if he also bought and borrowed it in some such worlds). Thus, the *exh* operator in (8) is not only motivated by the truth condition in (5b), but also to prevent trivialization of the modal component of *einfach*.

(12) $\text{Alt}(S_1) = \{[\lambda w\,.\,\text{H pirated } L \text{ in } w], [\lambda w\,.\,\text{H bought } L \text{ in } w], [\lambda w\,.\,\text{H borrowed } L \text{ in } w]\}$

(13) $\text{Alt}(S_2) = \{[\![\text{exh}]\!](\text{Alt}(S_1))(p) : p \in \text{Alt}(S_1)\}$
$= \{[\lambda w\,.\,\text{H pirated } L \text{ in } w \wedge \neg\text{H bought } L \text{ in } w \wedge \neg\text{H borrowed } L \text{ in } w], \quad (= [\![S_2]\!])$
$[\lambda w\,.\,\text{H bought } L \text{ in } w \wedge \neg\text{H pirated } L \text{ in } w \wedge \neg\text{H borrowed } L \text{ in } w], \ldots\}$

3 The Denotation of *einfach irgendein*

In line with our previous assumptions, we assume that (3) has the LF in (14).

(14) einfach_i [$_{S_2}$ exh [$_{S_1}$ [$\text{irgendein}_\text{D}$ Buch] [1 [Hans_i t_1 gekauft hat]]]]

We follow [3] in assuming that *irgendein* comes with a covert domain variable D and that the set assigned to D is contextually determined. Thus, S_1 has the denotation in (15), where $D^* = a(\text{D})$.

(15) $[\![S_1]\!]^a = [\lambda w\,.\,\exists x \in D^*[x \text{ is a book in } w \wedge \text{Hans bought } x \text{ in } w]]$

We proceed by noting that (3) gives rise to the exhaustivity inference that Hans didn't buy several books.[6] Importantly, this inference concerns all objects in the restriction and scope of *irgendein* and not just books that Hans bought randomly. To see this, assume that Hans went to his favorite bookstore and bought a random book for Marie and a carefully selected book for himself (and no other books). To report about this situation on the following day, only the variant of (16) that includes the phrase in parentheses can be adequately used.

(16) *Gestern hat Hans in seinem Lieblingsbuchladen einfach irgend-ein Buch #(für Marie) gekauft.*
yesterday has Hans in his favorite bookstore simply IRGEND-a book for Marie bought
'Yesterday, Hans simply bought a random book #(for Marie) at his favorite bookstore.'

[6] We thank Luka Crnič for making us aware of this inference and for discussing with us how to derive it.

To account for the observed exhaustivity inference, we follow [3] and [8] in assuming that the alternative set of *irgendein* ($\exists x \in D^*$) includes all of its subdomain alternatives ($\exists x \in D$, for all nonempty $D \subseteq D^*$), a universal alternative ($\forall x \in D^*$), and all subdomain alternatives of the universal alternative ($\forall x \in D$, for all nonempty $D \subseteq D^*$), as shown in (17).[7]

(17) $\begin{aligned}\text{Alt}(S_1) &= \{[\lambda w\,.\,\exists x \in D[x \text{ is a book in } w \wedge \text{Hans bought } x \text{ in } w]] : \varnothing \subset D \subseteq D^*\} \\ &\quad \cup \{[\lambda w\,.\,\forall x \in D[x \text{ is a book in } w \wedge \text{Hans bought } x \text{ in } w]] : \varnothing \subset D \subseteq D^*\} \\ &= \{[\text{H bought a book from } D] : \varnothing \subset D \subseteq D^*\} \\ &\quad \cup \{[\text{H bought every book from } D] : \varnothing \subset D \subseteq D^*\} \qquad \text{(abbrev.)}\end{aligned}$

Next, departing from [3], we assume that *exh* respects the innocent excludability of the alternatives in its domain ([10, 8]).[8] Hence, the denotation of S_2 entails that Hans didn't buy several books:[9]

(18) $\begin{aligned}[\![S_2]\!]^a &= [\lambda w\,.\,\exists x \in D^*[x \text{ is a book in } w \wedge \text{Hans bought } x \text{ in } w \\ &\qquad\qquad \wedge \neg\exists y \in D^*[x \neq y \wedge y \text{ is a book in } w \wedge \text{Hans bought } y \text{ in } w]]] \\ &= [\text{H bought a book from } D^* \text{ and no other book from } D^*] \qquad \text{(abbrev.)}\end{aligned}$

Furthermore, these assumptions yield that $\text{Alt}(S_2)$ is the set given in (19).

(19) $\begin{aligned}\text{Alt}(S_2) &= \{[\![\text{exh}]\!](\text{Alt}(S_1))(p) : p \in \text{Alt}(S_1)\} \\ &= \{[\lambda w\,.\,\exists x \in D[x \text{ is a book in } w \wedge \text{Hans bought } x \text{ in } w \\ &\qquad\qquad \wedge \neg\exists y \in D^*[x \neq y \wedge y \text{ is a book in } w \wedge \text{Hans bought } y \text{ in } w]]] : \varnothing \subset D \subseteq D^*\} \\ &\quad \cup \{[\lambda w\,.\,\forall x \in D[x \text{ is a book in } w \rightarrow \text{Hans bought } x \text{ in } w] \\ &\qquad\qquad \wedge \neg\exists x \in D^* \setminus D[x \text{ is a book in } w \wedge \text{Hans bought } x \text{ in } w]] : \varnothing \subset D \subseteq D^*\} \\ &= \{[\text{H bought a book from } D \text{ and no other book from } D^*] : \varnothing \subset D \subseteq D^*\} \\ &\quad \cup \{[\text{H bought every book from } D \text{ and no book from } D^* \setminus D] : \varnothing \subset D \subseteq D^*\} \qquad \text{(abbrev.)}\end{aligned}$

Henceforth, we ignore the universal alternatives in $\text{Alt}(S_2)$ since they are irrelevant for the validity of the arguments that follow. Thus, given the semantics of *einfach* in §2, we end up with the following truth conditions for (3):

(20) a. Hans bought a book from D^* and no other book from D^*.
(*actuality & exhaustivity inference*)

b. For every nonempty $D \subseteq D^*$, there is a possible world, compatible with the circumstances of the actual world, in which Hans buys a book from D and no other book from D^*. (*circumstantial possibly inference*)

c. There is no $D \subseteq D^*$ such that [H bought a book from D and no other book from D^*] is simpler for Hans than [H bought a book from D^* and no other book from D^*].
(*simplicity inference*)

The actuality and exhaustivity inferences thus follow without further ado. What remains to be shown is that the circumstantial possibility inference and the simplicity inference effectively equate to (or entail) AI.

[7] According to [3], *irgendein* induces (i) subdomain alternatives by its lexical specification, and (ii) a universal alternative by being an indefinite. From [8], we can deduce the assumption that the alternative generation mechanism yields the Cartesian product of (i) and (ii).

[8] See §5, where we discuss this crucial departure from [3].

[9] To see this, note that none of the existential alternatives in (17) are innocently excludable, and neither are any of the universal alternatives with a singleton domain. However, all of the universal alternatives with a non-singleton domain are innocently excludable, which leads to the inference that Hans didn't buy several books.

4 Deriving Indifference

We now argue that, if an agent has options for action (e.g. buying this (kind of) book or that) and preferences about which option to realize, then realizing one of many options is more complex than realizing one of fewer options — intuitively, since realizing one of many options requires considering more options. From the truth of (3), we can infer that Hans has more book buying options if he's buying a book from D^* than if he's buying a book from $D \subset D^*$, since the circumstantial possibility inference (20b) entails that each book in D^* is buyable for Hans. Since, furthermore, the simplicity inference (20c) entails that Hans buying a book from D^* (many options) is no more complex than Hans buying a book from $D \subset D^*$ (fewer options), it follows that Hans has no preference about which (kind of) book to buy (*indifference inference*).

We continue using '$\text{Alt}(S_2)$' to refer to the domain of alternatives of *einfach* as given in (19) and proceed in two steps: (I) we show that the propositions in $\text{Alt}(S_2)$ can be associated with decision problems; (II) we show that book buying preferences have an impact on the complexity of these decision problems: for every (nonempty) $D \subset D^*$, the decision problem for [H bought a book from D and no other book from D^*] is simpler for Hans than the decision problem for [H bought a book from D^* and no other book from D^*] iff Hans has book buying preferences.

Step I. Assume that $k(x, w)$ is an ordering source that characterizes x's preferences in w and that $>_{k(x,w)}$ is the corresponding (strict) ordering relation between propositions. For example, assume that Hans has book buying preferences that lead to the orderings in (21) and to no other orderings of logically independent propositions (where $b_1, \ldots, b_4$ are four arbitrary books from D^*).

(21) a. [H bought b_1 and no other book from D^*]
$>_{k(\text{Hans},w)}$ [H bought b_2 and no other book from D^*]

b. [H bought b_3 and no other book from D^*]
$>_{k(\text{Hans},w)}$ [H bought b_4 and no other book from D^*]

Let $\text{Alt}(S_2)_{\Rightarrow p}$ be the set $\{q \in \text{Alt}(S_2) : q \Rightarrow p\}$. Then every $p \in \text{Alt}(S_2)$ defines a decision problem relative to $\text{Alt}(S_2)_{\Rightarrow p}$, namely the problem of identifying the weakest propositions $q \in \text{Alt}(S_2)_{\Rightarrow p}$ such that $\neg\exists r \in \text{Alt}(S_2)_{\Rightarrow p}$ with $r >_{k(\text{Hans},w)} q$. This problem corresponds to the problem of identifying the maximal subsets E of D^* (or one of its subsets) that satisfy Hans's book buying preferences in w no worse than any other subset (e.g. the problem of identifying the subset $\{b_1, b_3\}$ of $\{b_1, \ldots, b_4\}$ given (21)). To see this, consider the decision problem for

$$p_{\{b_1,\ldots,b_4\}} = [\text{H bought a book from } \{b_1, \ldots, b_4\} \text{ and no other book from } D^*]$$

relative to $\text{Alt}(S_2)_{\Rightarrow p_{\{b_1,\ldots,b_4\}}}$ and the preference ordering in (21). We will show that

$$p_{\{b_1,b_3\}} = [\text{H bought a book from } \{b_1, b_3\} \text{ and no other book from } D^*]$$

is the solution to the decision problem for $p_{\{b_1,\ldots,b_4\}}$. First, consider the alternative $p_{\{b_1\}}$. We note that it's not the case that $p_{\{b_1\}} >_{k(\text{Hans},w)} p_{\{b_1,b_3\}}$: $p_{\{b_1,b_3\}}$-worlds in which Hans bought b_3 are unordered relative to worlds in which he bought b_1 and no other book from D^*. Since, furthermore, $p_{\{b_1,b_3\}}$ is weaker than $p_{\{b_1\}}$, $p_{\{b_1\}}$ is not the solution to the decision problem for $p_{\{b_1,\ldots,b_4\}}$. By the same reasoning, $p_{\{b_3\}}$ is not the solution to the decision problem for $p_{\{b_1,\ldots,b_4\}}$, either. Next, consider $p_{\{b_1,b_2,b_3\}}$. We find that $p_{\{b_1,b_3\}} >_{k(\text{Hans},w)} p_{\{b_1,b_2,b_3\}}$ since worlds in which Hans bought b_2 and no other book from D^* are less preferred than $p_{\{b_1,b_3\}}$-worlds in which Hans bought b_1, and unordered relative to $p_{\{b_1,b_3\}}$-worlds in which Hans bought b_3. By the same reasoning, $p_{\{b_1,b_3,b_4\}}$ and $p_{\{b_1,\ldots,b_4\}}$ are less preferred than $p_{\{b_1,b_3\}}$, too. Thus, $p_{\{b_1,b_3\}}$ is the solution to the decision problem for $p_{\{b_1,\ldots,b_4\}}$.

Step II. If Hans has book buying preferences, then for all (nonempty) sets $D \subset D^*$ the decision problem for p_D = [H bought a book from D and no other book from D^*] is simpler than that for p_{D^*} = [H bought a book from D^* and no other book from D^*]: $\text{Alt}(S_2)_{\Rightarrow p_D}$ is a proper subset of $\text{Alt}(S_2)_{\Rightarrow p_{D^*}}$, since p_D asymmetrically entails p_{D^*}; consequently, the decision problem for p_D relative to the former set requires considering less alternatives than the decision problem for p_{D^*} relative to the latter set. If Hans has no book buying preferences, then the decision problem is trivial for all $D \subseteq D^*$: the proposition sought after is [H bought a book from D and no other book from D^*].

Putting everything together. We assume that the complexity of the decision problems associated with the members of $\text{Alt}(S_2)$ determines how simple the members of $\text{Alt}(S_2)$ are for Hans: for all $p, q \in \text{Alt}(S_2)$, p is simpler for Hans than q iff the decision problem for p is simpler for Hans than the decision problem for q. Then, (I) and (II) show that the simplicity inference (20c), in conjunction with the circumstantial possibility inference (20b), entails that Hans didn't have book buying preferences, and hence that (3) entails that Hans was indifferent about the type and specimen of book he bought.

5 Discussion

We end with a discussion of how our analysis captures several interesting differences between German *irgend-*, on the one hand, and Spanish *cualquiera* and English *any*, on the other hand. We also describe a new puzzle arising from our proposal that AI is the result of an interaction between a modal operator (*einfach*) and subdomain alternatives.

Comparison with Spanish *cualquiera*. As we mentioned in §1, AOMB argue that Spanish *un NP cualquiera*, which, like *irgendein NP*, triggers an AI reading (cf. (2)), is best analyzed as having a modal component hardwired into the meaning of *cualquiera*. Their motivation is that the AI reading easily persists even when the indefinite occurs in a downward-entailing (DE) environment, as in (22), which would be unexpected if AI were merely conversationally implicated, for instance.

(22) *Juan no compró un libro cualquiera para María.*
Juan not bought a book CUALQUIERA for María
'Juan didn't buy a random book for María.'

On our proposal for German, by contrast, AI arises via the interaction of *irgend-*, which triggers subdomain alternatives, and *einfach*, which may have *exh* in its immediate scope. Consider now (23), in which *irgendein Buch* occurs in the scope of the DE operator *nie* 'never'. Our proposal predicts that, without any *einfach*, (23) simply means that Hans didn't buy any book, and indeed this a natural reading of the sentence (see [3]). If, however, *einfach* (overt or otherwise) is inserted, then, in order to avoid a contradiction, *exh* must also occur in its scope (hence, in the scope of *nie*).[10] It's well known, however, that the distribution of *exh* is rather limited, in particular that it isn't happy in DE contexts, unless special stress is added to the item that triggers the alternatives in the domain of *exh* (see, e.g., [11]). As such, we predict that embedded AI readings of *(einfach) irgend-* can occur in German, but only if special stress is added to the indefinite, and this appears to be exactly right (cf. [13]).

[10] Recall from §2 that without *exh*, the modal component of *einfach* is trivially satisfied. As such, in DE contexts, without *exh*, the reverse occurs; namely, the modal component is contradictory (unsatisfiable). The same prediction arises if we assume that exhaustification is part of the meaning of *einfach* itself.

(23) *Hans hat nie irgend-ein Buch gekauft.*
Hans has never IRGEND-a book bought
With stress on *irgendein*: 'Hans has never bought a random book.'
Without stress on *irgendein*: 'Hans has never bought any book.'

Comparison with English *any*. In §3, we derived the AI reading of (3) from what we assumed to be the LF structure underlying this reading, viz. (14). Assuming this LF structure, however, is not innocuous since (14) contains a substructure that has a peculiar status in the theory of polarity sensitivity of [3]. The substructure in question is the complement of *einfach*, which is of the form in (24). Recall that by our assumptions *irgendein* is an indefinite that activates subdomain alternatives and that these subdomain alternatives are contained in the domain of *exh*. Moreover, the complement of *exh* is an upward-entailing (UE) environment for *irgendein*, as indicated by the subscript.

(24) exh $[_{\text{UE}} \ldots \text{irgendein} \ldots]$, where *exh* ranges over the subdomain alternatives of *irgendein*

What is peculiar about the structure in (24) is that it denotes the contradiction if, as is assumed in [3], *exh* doesn't respect the condition of innocent excludability of the alternatives in its domain. Since we assume, in contrast, that *exh* does respect this condition (see §3), we derive a contingent proposition from (24) which, together with its alternative propositions and the meaning of *einfach*, entails AI. Thus, we disagree with [3] on the polarity sensitivity of *irgendein*, in particular, and on the definition of *exh* and, hence, its role in explaining the distribution of polarity sensitive items, in general. As for the former disagreement, we note that, unlike English *any*, *irgendein* can occur in what appears to be an unembedded position in a plain declarative sentence, as illustrated by (25a) vs. (25b). [3] takes the modal implicature triggered by *irgendein* in such sentences (see the paraphrase of (25b)) to show that *irgendein* is separated from *exh* by a covert modal operator (which prevents a contradictory meaning from emerging). That is, [3] assumes that (25b) has an LF structure of the form *exh* $[\Diamond [\textit{irgendwer} \ldots]]$, where $\Diamond$ is a covert modal operator. We submit that at least the reading of (25b) on which it implicates speaker ignorance does not provide evidence for a covert modal. Rather, the speaker ignorance implicature follows straightforwardly from the Gricean maxim of quantity if (25b) has the form *exh* $[\textit{irgendwer} \ldots]$ (where *exh* is the operator of [10], which respects innocent excludability) and *irgend-* activates subdomain alternatives as assumed in §3 (following [3]).[11] That is, we hold that the best explanation for the paradigm in (25) and (26) is that English *any* differs from German *irgend-* and from the English and German disjunctive particles in that it is a polarity sensitive item, while the other items, which trigger speaker ignorance inferences, are not.

(25) a. *Anyone called.
b. *Irgend-wer hat angerufen.*
IRGEND-who has called
'Someone called (and the speaker doesn't know or care to tell who).'

(26) a. Ann or Bill called.
b. *Anne oder Willi hat angerufen.*
Anne or Willi has called
'Anne or Willi called (and the speaker doesn't know which one of the two).'

We are not yet in a position to say if our analysis of the AI reading of *einfach ... irgend-* is compatible with any of the existing theories of polarity sensitivity.[12] However, provided that

[11] If, alternatively, the speaker ignorance reading of (25b) is caused by a syntactically represented modal operator as argued in [14], *irgendwer* may still be immediately c-commanded by an occurrence of *exh*. If we follow [14], we are led to assume that the relevant reading of (25b) is due to an LF structure of the form *exh* $[K$ $[\textit{exh}$ $[\textit{irgendwer} \ldots]]]$, where the lower occurrence of *exh* ranges over the subdomain alternatives of *irgendwer*.

[12] The theory defended in [7, 6], which is not based on subdomain alternatives being associated with *exh* but rather with a covert variant of *even* ([12]), may be a suitable candidate.

such a theory exists, our analysis makes the following prediction: if English has a counterpart of German *einfach*, then *any* can give rise to AI readings in environments in which it can occur as a polarity sensitive item, e.g. in the immediate scope of a sentence negation. We submit that English *just* is the relevant counterpart of *einfach* and that our prediction is borne out (if the proviso can be satisfied): the sentence in (27a) has a reading on which it implies AI (where small capitals indicate that *any* must be stressed for the AI reading to arise, for reasons discussed in the previous subsection). Furthermore, there is evidence that *just*, like its German counterpart, has a covert variant, as is evidenced by the sentence in (27b), which has an AI reading.[13]

(27) a. John didn't buy just ANY book.
b. Don't buy ANY data plan. (Buy ours!)

New puzzle: Disjunction and the lack of agent indifference. There is an intuitively close connection between indefinites and disjunction, in the sense that a sentence with an indefinite can be thought of as disjunctive in meaning: if the set of all (relevant) books is just $\{Faust, Lolita\}$, then *Hans bought a book* is semantically equivalent to *Hans bought* Faust *or* Lolita. Within semantic theory, it's also common to assume that disjunctions, like indefinites, trigger (what we might call) subdomain alternatives: the alternatives of *Hans bought* Faust *or* Lolita include not just the conjunctive alternative *Hans bought* Faust *and* Lolita, but also the individual disjunct alternatives, *Hans bought* Faust and *Hans bought* Lolita (see [15]). If this is correct, however, then we appear to predict that disjunctive sentences can have AI readings: Hans arbitrarily bought one of *Faust* or *Lolita*, without any preference. Unfortunately, this prediction is not borne out, as neither the English sentence nor its German equivalent (with or without overt *einfach*) can be understood in that way. That being said, we stress that this appears to be a general puzzle that arises for any straightforward account of the AI effects of *irgend-*, together with standard assumptions about subdomain alternatives: whatever mechanism results in universal inferences about subdomain alternatives for *irgend-* seems to likewise result in universal inferences about sub-disjunction alternatives for plain disjunctions. We of course must leave a solution to this puzzle for a future occasion.

6 Conclusion

The German indefinite modifier *irgend-* can license the inference that the agent of an action was in some sense indifferent as to the outcome of the action. We proposed a novel and intuitive analysis of agent indifference by building on well-established assumptions about the semantics of indefinites and on new observations about the role of the adverbial *einfach* 'simply'. *Irgend-* activates subdomain alternatives, while *einfach* licenses a simplicity inference. In *einfach irgend-* sentences, the simplicity inference is, roughly, that doing an action relative to a large domain D^* is no more complex for the agent than doing that action relative to a subdomain D, and this, we argued, can only be the case if the agent has no preferences about the outcome of the action, i.e. is indifferent. Our proposal correctly predicts that AI readings of *irgend-* embedded in a DE context can arise, but only if the indefinite is stressed, hence captures an important difference between *irgend-* and Spanish *cualquiera*. In addition, to the extent that our proposal can be

[13] An anonymous reviewer pointed out to us that for them sentences like in (27b) cannot imply AI (though we aren't sure whether the reviewer controlled for stress). We are confident that our empirical claim is correct for at least some speakers of English, since our example is a simplified version of an an actual advertisement that is meant to convince listeners to buy a mobile data plan in a non-random way, and not to refrain from buying a data plan altogether. More to the point, if the first sentence in (27b) couldn't imply AI, then the sequence as a whole would sound contradictory, and yet it doesn't.

supplemented by a theory of the polarity sensitivity differences between *irgend-* and English *any*, it correctly predicts that AI readings of *any* can arise, but only in DE contexts, hence captures an important difference between *irgend-* and English *any*.

While we find our account to be both intuitive and plausible, we've only sketched a proof-of-concept of how the simplicity relation over propositional alternatives that *einfach* references can yield agent indifference — namely, by assuming that it's determined by associated decision problems. A fully explicit theory needs to not only make this link precise, but also explain why the simplicity order can't be provided by some other metric than the one suggested here.

References

[1] Luis Alonso-Ovalle and Paula Menéndez-Benito. Expressing indifference: Spanish *un* NP *cualquiera*. In *Semantics and Linguistic Theory (SALT)*, volume 21, pages 333–352, 2011.

[2] Luis Alonso-Ovalle and Paula Menéndez-Benito. Projecting possibilities in the nominal domain: Spanish *uno cualquiera*. *Journal of Semantics*, 2017. Forthcoming.

[3] Gennaro Chierchia. *Logic in Grammar: Polarity, Free Choice, and Intervention*. Oxford University Press, Oxford, England, 2013.

[4] Gennaro Chierchia, Danny Fox, and Benjamin Spector. Scalar implicature as a grammatical phenomenon. In Claudia Maienborn, Klaus von Heusinger, and Paul Portner, editors, *Semantics: An International Handbook of Natural Language Meaning*, volume 3, pages 2297–2331. Mouton de Gruyter, Berlin, Germany, 2012.

[5] Jinyoung Choi. *Free Choice and Negative Polarity: A Compositional Analysis of Korean Polarity Sensitive Items*. PhD thesis, University of Pennsylvania, 2007.

[6] Luka Crnič. Against a dogma on NPI licensing. In Luka Crnič and Uli Sauerland, editors, *The Art and Craft of Semantics: A Festschrift for Irene Heim*, volume 1, pages 117–145. MIT Working Papers in Linguistics, Cambridge, MA, 2014.

[7] Luka Crnič. Non-monotonicity in NPI licensing. *Natural Language Semantics*, 22(2):169–217, 2014.

[8] Luka Crnič. Free choice under ellipsis. *The Linguistic Review*, 34(2):249–294, 2017.

[9] Kai von Fintel. *Whatever*. In *Semantics and Linguistic Theory (SALT)*, volume 10, pages 27–39, 2000.

[10] Danny Fox. Free choice and the theory of scalar implicatures. In Uli Sauerland and Penka Stateva, editors, *Presupposition and Implicature in Compositional Semantics*, Palgrave Studies in Pragmatics, Language and Cognition Series, chapter 4, pages 71–120. Palgrave Macmillan, New York, NY, 2007.

[11] Danny Fox and Benjamin Spector. Economy and embedded exhaustification. *Natural Language Semantics*, 2017. Forthcoming.

[12] Irene Heim. A note on negative polarity and downward entailingness. In *North East Linguistic Society (NELS)*, volume 14, pages 98–107, 1984.

[13] Angelika Kratzer and Junko Shimoyama. Indeterminate pronouns. In Yukio Otsu, editor, *Tokyo Conference on Psycholinguistics*, volume 3, pages 1–25, Tokyo, Japan, 2002. Hituzi Syobo.

[14] Marie-Christine Meyer. *Ignorance and Grammar*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, 2013.

[15] Uli Sauerland. Scalar implicatures in complex sentences. *Linguistics and Philosophy*, 27(3):367–391, 2004.

[16] Elisabeth Villalta. Mood and gradability: An investigation of the subjunctive mood in Spanish. *Linguistics and Philosophy*, 31:467–522, 2008.

Plurality in Buriat and structurally constrained alternatives *

Lisa Bylinina[1] and Alexander Podbrjaev[2]

[1] Leiden University,
Leiden, the Netherlands
[2] Higher School of Economics,
Moscow, Russia

Abstract

In this paper we offer a solution to a puzzle in the number interpretation of nominals in Buriat. Buriat has a two-way number opposition in morphology (unmarked vs. plural), but semantically, both forms may be number neutral. We show that even though the number neutrality of unmarked nominals is heavily restricted (to inanimate nouns), it does not boil down to incorporation or pseudo-incorporation. Our proposal is that unmarked nominals can be either singular (projecting a NumP) or numberless (lacking a NumP). In case they are singular, they are semantically strictly atomic, but when there are numberless they are truly number neutral, just like the plurals. The plurality inferences of plurals and the consistent number neutrality of numberless nouns are accounted for in a Katzirian system with structurally defined alternatives.

1 The Puzzle

Nouns in Barguzin dialect of Buriat as spoken in the village of Baragkhan, Republic of Buryatia, Russian Federation (henceforth referred to simply as Buriat) show morphological distinction between two forms: one traditionally referred to as 'singular' (morphologically unmarked) – and 'plural' (hosting an overt plural suffix):[1]

(1) a. nom 'book' vs. nom-uːd 'books'
b. xʉbʉː(n) 'boy' vs. xʉbʉː-d 'boys'

In this paper, we focus on the range of number interpretations of morphologically unmarked and morphologically plural forms in different contexts. First, we show that the interpretation of these forms seems to posit a problem for two major classes of semantic theories of number, for which we use the labels STRONG SG / WEAK PL theory and WEAK SG / STRONG PL theory. Then, we introduce further data that will help us resolve the problem in favour of the STRONG

*The data discussed in this paper was collected during a field trip to Baragkhan village, Kurumkansky District, Republic of Buryatia, Russian Federation, during the summer of 2017. The authors thank our language consultants, as well as the Department of Theoretical and Applied Linguistics of Moscow State University for organizing this trip and letting us participate in it. The paper emerged as an outcome of collaboration between the first author, whose work on number and plurality is supported by a grant from the Netherlands Organisation for Scientific Research / VENI Grant no. 275-70-045, and the second author, whose study of syntax and semantics of Buriat has been conducted at Lomonosov Moscow State University as part of the project #16-18-02081 funded by the Russian Science Foundation.

[1] Buriat has several plural morphemes, each comes with non-trivial morphophonological properties. For the purposes of the current paper, we will treat them as variants of one plural suffix due to the lack of semantic differences between them. We also don't discuss stem alternations involving final *n* that will force the words like *xʉbʉː(n)* 'boy' to appear with or without it in different envirnonments.

SG / WEAK PL theory in combination with structural constraints on alternatives (as described in Katzir 2007; Fox and Katzir 2011).

Let's start with the unmarked ('singular') form. In Buriat, inanimate nouns unmarked for plurality systematically get number-neutral interpretation in a range of contexts, illustrated here for the direct object position:

(2) bʲi nom unʃ-aː-b
I book read-PST-1SG
'I read a book / books'

The word *nom* 'book' in (2) doesn't have number marking (or any other marking, for that matter) and, in this sentence, it can refer to one book or to more than one book.

This number interpretation of morphologically unmarked inanimate nouns is not restricted to direct object positions – the possibilities include (genitive) object of a postposition (3) and nominative subject (4):

(3) bʲi nom-iːn tʉləː xoʒomdoː-b
I book-GEN because.of was.late-1SG
'I was late because of the book / books'

(4) nom honʲin baig-aː
book interesting be-PST
'The book(s) was/were interesting'

They can also be subjects of collective predicates – predicates that require objects in their denotation to be pluralities:

(5) nom olon baig-aː
book many be-PST
'There were many books' / 'The books were many'

Morphologically unmarked *animate* nouns don't give rise to number-neutral interpretation – (6), for example, is only compatible with the speaker having seen one boy:

(6) bʲi xʉbʉː xar-aː-b
I boy see-PST-1SG
'I saw a boy / #boys'

Morphologically plural nouns (both inanimate and animate) in Buriat give rise to non-singularity inferences in upward-entailing (UE) contexts – (7) requires there to be more than one book the speaker was late because of; in (8) there was strictly more than one book that was interesting:

(7) bʲi nom-uːd-iːn tʉləː xoʒomdoː-b
I book-PL-GEN because.of was.late-1SG
'I was late because of the books / #book'

(8) nom-uːd honʲin baig-aː
book-PL interesting be-PST
'The books were / #book was interesting'

In downward-entailing (DE) contexts, however, the non-singularity inferences of morphologically plural nouns disappear – (9) is false if the speaker has one Buriat book; the question in

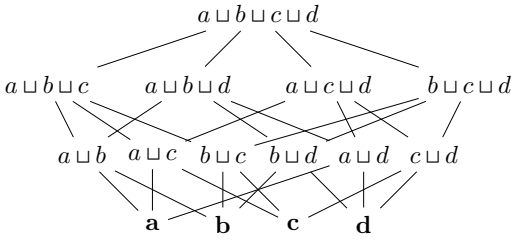


Figure 1: The domain of entities

(10) can get a true positive answer in case the addressee has only read one book in Buriat; in (11), one book satisfies the condition:

(9) namda burʲaːd nom-uːd ʉgiː
I.DAT Buriat book-PL COP.NEG
'I don't have Buriat books'

(10) ʃi xɘzɘːʃtaː burʲaːdaːr nom-uːd-iːjɘ unʃ-aː-nʃ?
you ever Buriat-INST book-PL-ACC read-PST-2SG
'Have you ever read books in Buriat?'

(11) ʃamda burʲaːd nom-uːd biː xada-nj, tɘdɘː-niːɘ asar-aa-raj
you.DAT Buriat book-PL COP if-3.POSS they-ACC bring-PST-PRSCR
'If you have Buriat books (even if you only have one), bring them.'

For most of this paper, we will focus on inanimate nouns, suggesting a speculation about number on animate nouns towards the end of the paper. For inanimate nouns in Buriat the following generalizations arise given the facts presented above: 1) nouns morphologically unmarked for number are semantically number-neutral (their denotation includes both singularities and pluralities); 2) morphologically plural nouns are semantically number-neutral as well (as revealed by DE contexts).

To make our reasoning easier, let's formulate these generalizations against a formal background in which the domain of individuals has the structure of complete join semilattice (Link 1983; Landman 1991, 2000 a.o.), see figure 1. The structure of the domain captures the 'part-of' relation – say, that John is part of John and Bill. Under this approach, there is no type-theoretic difference between singular and plural individuals, plural individuals are type e entities just like singular ones. The distinction between atomic (*john*, *bill*, ...) and sum (*john*⊔*bill*, ...) subdomains will be crucial. We notate predicates ranging exclusively over atoms as P, predicates with the whole semilattice as their domain will be ${}^*P$ (* being closure of P under 'join'), and predicates ranging over the non-atomic part of the semilattice will be ${}^*P \backslash P$.

Reformulating our generalization in these terms, both morphologically unmarked and morphologically plural Buriat nouns are ${}^*P$ predicates.

The puzzle is thus twofold: How come two forms with different number marking have the same number interpretation – ${}^*P$? And under what mechanism does one of these forms comes to bear non-singularity inferences?

Before moving on to approach this puzzle, we will show that the situation in Buriat is different from those described for other languages with with semantically number-neutral morphologically unmarked nouns.

2 Types of languages with number-neutrality

Buriat is far from the only known language in which nouns unmarked for number exhibit semantic number-neutrality. Similar observations have been made at least for Turkish (Öztürk, 2005; Bale et al., 2010), Hungarian (Farkas and de Swart, 2010), Western Armenian (Bale and Khanjian, 2014) and Hindi (Dayal, 2011). However, Buriat is different from all of these languages – either with respect to the properties of the unmarked forms, or with respect to the properties of the plural ones.

In most languages in question the distribution of number neutral unmarked nominal forms is very limited. In Hindi (Dayal, 2011), Hungarian (Farkas and de Swart, 2003) or Turkish (Öztürk, 2005), they can only be found in the (pseudo-)incorporation construction. It has been argued that such constructions don't involve full-fledged DPs, and these forms don't have argumental semantic type at all. This DP deficiency has been linked to number-neutrality (Farkas and de Swart 2003 a.o.).

But in Buriat, it would be hardly possible to treat all instances of number neutrality of unmarked forms as cases of pseudo-incorporation. The range of syntactic positions where number neutrality arises in Buriat is greater than what (pseudo-)incorporation is usually assumed to be able to target (although see Öztürk 2005 on the possibility of subject pseudo-incorporation in Turkish). Even for direct objects (DOs), which would be most likely to undergo (pseudo-)incorporation, it is easy to demonstrate that they lack the hallmark properties of the construction, which suggests that they are full-fledged DPs: they are separable from their predicate, can serve as antecedents of discourse anaphora, don't interact with aspect in a way typical for pseudo-incorporation and can have wide scope w.r.t. other quantificational elements in the sentence (tests following Farkas and de Swart 2003; Mithun 2010 a.o.). Here we illustrate the scopal behaviour of unmarked forms and their interaction with aspect.

(pseudo-)incorporation normally comes with obligatorily narrow scope, however, in (12), number-neutral *nom* can have wide scope with respect to the modal. (12) can be truthfully used in a situation in which: 1) there is a requirement to buy more than one book; 2) the quantity and identity (say, titles) of the books are part of the requirement:

(12) bʲi nom xudalda-ʒa aba-xa johotoj-b
I book sell-CONV take-POT need-1SG
'I have to buy a book / books'
(Can be used if the requirement is to buy 'War and Peace' and 'Crime and Punishment')

Dayal (2011) argues that NPs undergoing pseudo-incorporation in Hindi and Hungarian are always specified for number (singular), and the apparent number neutrality is a result of embedding under certain aspectual operators. Crucially for her argument, pseudo-incorporated DOs in Hindi are incompatible with the telic interpetation. In Buriat this generalization does not seem to hold. In (13), the unmarked *nom* 'book' denotes a plurality in a telic clause.

(13) ʉglɵː-gʉːr xʉbʉːn nom ʉnʃ-aː-d bai-gaː
morning-INST boy book read-CONV be-PST
'By the morning, the boy has read the books'

Beyond that, there is extensive evidence that number neutrality of unmarked nouns in Buriat cannot be reduced to atomicity under aspectual operators. Consider (14), which is ambiguous between an atomic and a non-atomic interpretation of the unmarked noun, with the non-atomic interpretation (*hiding between houses*) clearly **not** arising from quantifying over events of *hiding in the middle of a house*, even under an atelic interpretation.

(14) badma gɔr dunda xorgod-oo
Badma house middle hide-PST
a. 'Badma hid in the middle of a house'
b. 'Badma hid between houses'

We conclude that unmarked number-neutral nominal forms in Buriat are not (pseudo-)incorporated, and can constitute regular full-fledged argumental DPs, semantic number-neutrality thus being a property of unmarked argumental DPs in Buriat.

Western Armenian may be the only language that has been claimed to have number-neutral NPs unmarked for number in argument positions without (pseudo-)incorporation (Bale and Khanjian, 2014). In this respect, Western Armenian patterns with Buriat, however, there is a crucial difference. As Bale and Khanjian (2014) argue, plural forms in Western Armenian are not semantically number-neutral, but rather range over only the non-atomic part of the domain of entities ($^*P \backslash P$). In Buriat, on the other hand, there is evidence that the plural forms are number neutral. The number neutrality of plurals becomes evident in non-upward-entailing environments, as in examples (9)-(11) above.

It seems that situation in Buriat is unique: it has a two-way number distinction in morphology (unmarked and plural), both forms show semantic number-neutrality in argument positions – unmarked number-neutral forms can't be brushed off as pseudo-incorporation, plural forms can't be argued to denote strictly non-atoms. The next section briefly summarizes two big classes of theories of number as candidates for an analysis for this situation.

3 Two theories of number

Semantic theories of number can be divided into two broad groups – after (Farkas and de Swart, 2010), we call them STRONG SG / WEAK PL theories and WEAK SG / STRONG PL theories.

STRONG SG / WEAK PL theories of number offer an analysis for the situation in which morphologically plural forms get number-neutral ($^*P$) interpretation in DE contexts and strictly plural ($^*P \backslash P$) interpretation in UE contexts (the situation in Buriat). According to these theories, morphologically plural forms have both atoms and sums as their domain ($=^*P$) (Sauerland, 2003; Sauerland et al., 2005; Spector, 2007; Zweig, 2009). To account for non-singular inferences in UE contexts, these theories invoke pragmatic mechanisms relying on the singular form as an alternative to the plural form (P vs. $^*P$ semantically). Implementations range from Maximize Presupposition (Sauerland, 2003; Sauerland et al., 2005) – to scalar implicature based on comparative logical strength of P and $^*P$ alternatives in context (Spector 2007 a.o.). Importantly, all these theories build on singular alternative being atom-denoting (P). This is where Buriat data becomes problematic – in the examples we've seen so far, morphologically unmarked forms don't have exclusively atomic reference. Similar concerns have been raised for non-singularity inferences of Japanese plurals in (Sudo, 2017).

Under one version of WEAK SG / STRONG PL approach (Bale and Khanjian, 2014), the domain of singular nouns includes both atoms in sums ($^*P$) and plural forms are strictly plural ($^*P \backslash P$). Singular forms sometimes – but not always – give rise to non-plurality inferences. They do so when in proper competition with morphologically plural forms. The conditions for such competition are structural: (Bale and Khanjian, 2014) argue that in Western Armenian, unmarked indefinites are not embedded in a DP, while indefinites marked for plural are, and this makes them too different structurally for competition, so the non-plurality inferences of unmarked forms don't arise. Unmarked and plural definites, to the contrary, both form DPs and thus compete, which gives rise to non-plurality inferences of unmarked forms.

Although we believe that structural properties are crucial for number inferences (see below), two considerations preclude application of this particular theory to Buriat data: 1) Unmarked number-neutral forms *do* form DPs in Buriat, unlike what (Bale and Khanjian, 2014) argue for Western Armenian; 2) Buriat plural forms are not strictly non-atomic but number-neutral.

A more complicated version of the WEAK SG / STRONG PL approach (Farkas and de Swart, 2010) suggests that the domain of singular nouns (in Hungarian, and potentially more generally) includes both atoms and sums ($=^*P$), while plural forms are ambiguous between the same and exclusively non-atomic reading ($^*P \backslash P$). This seems promising as it in principle allows for both unmarked and plural forms to satisfy the diagnostics for semantic neutrality ($^*P$), as is the case in Buriat. According to (Farkas and de Swart, 2010), plurals are subject to an additional requirement of having sum witnesses in their denotation, precluding them from having exclusively atomic reference. The choice between the $^*P$ and $^*P \backslash P$ readings of the plural is regulated by a pragmatic principle (Strongest Meaning Hypothesis), giving rise to non-singularity inferences in UE contexts. Singular DPs under this view are structurally strictly simpler than plural DPs – they lack a layer hosting the privative [Pl] feature. Still, in argument positions singulars and plurals form alternatives, and via this competition a strictly singular reading of non-plurals should arise. This is compatible with Hungarian data – non-pseudo-incorporated argumental unmarked forms semantically are strictly atomic. However, this is not the situation in Buriat, as shown above – argumental unmarked DPs are still semantically number-neutral. Apart from this empirical problem, this account has a theoretical problem – derivation of the non-plurality inference for the unmarked form invokes an alternative that is structurally more complex than the original item.

Summing up, existing theories don't cover Buriat data. Either they rely on the basic meanings that cannot be maintained for Buriat, or make wrong empirical predictions, while being problematic in the light of what is known about structural constraints on alternatives. The next section section explicates such constraints, relying on (Katzir, 2007; Fox and Katzir, 2011), and introduces more Buriat data that strengthens the point that structural considerations are relevant for number interpretation in Buriat. After that, we can formulate our analysis.

4 Structural constraints

As discussed above, accounts of number inferences of DPs often make use of some mechanism that crucially refers to the set of alternatives of a nominal form.

This section discusses one constraint on the set of alternatives, the constraint that has been argued to be active no matter what particular mechanism using alternatives this set is then input to (scalar implicature, focus, etc.). (Katzir, 2007; Fox and Katzir, 2011) argue that the ability of a structure to enter the set of alternatives of some other structure depends on the relative complexity of these structures. Here is how structural complexity is defined:

(15) STRUCTURAL COMPLEXITY (somewhat simplified)
Let ϕ, ψ be parse trees. If we can transform ϕ into ψ by a finite series of deletions, contractions, and replacements of constituents in ϕ with constituents of the same category taken from the lexicon, we will write $\psi \lesssim \phi$. If $\psi \lesssim \phi$ and $\phi \lesssim \psi$, we will write $\psi \sim \phi$.

For a structure ϕ, the alternatives will be all those structures that are at most as complex as ϕ:

(16) STRUCTURALLY DEFINED ALTERNATIVES
Let ϕ be a parse tree. The set of structural alternatives for ϕ, written as A_{str}, is defined as $A_{str}(\phi) := \{\phi' \mid \phi' \lesssim \phi\}$

Motivating examples for this kind of constraint are along the lines of (17) (from Katzir 2007):

(17) a. If any **tall man** comes to the party, he will be disappointed.
b. If any **man** comes to the party, he will be disappointed.

(17-b) $\lesssim$ (17-a). DE context makes sure that the less complex pair is the more informative one. Empirically, (17-a) bears an implicature that the (17-b) couldn't have been used instead. This means that (17-b) $\in A_{str}$(17-a), although it is strictly less complex than (17-a) – we can transform (17-a) into (17-b) by deletion.

If the context is changed from DE to UE, the more complex structure will be the more informative one. However, empirically, (18-a) does not have an implicature that (18-b) was not assertable – in fact, (18-a) suggests nothing about its (18-b) counterpart, so (18-b) $\notin A_{str}$(18-a):

(18) a. A **man** came to every party.
b. A **tall man** came to every party.

Back to semantics of number in Buriat – can unmarked and plural forms in Buriat end up in each other's A_{str}? Would non-singularity inferences of plural forms be then derivable? What is the structure of the DPs these forms are part of?

To answer these questions, we introduce further data showing that number readings of non-plural forms in Buriat are conditioned morphosyntactically. Configurations that induce strictly atomic reference include DOs with overt Acc case marking, 1&2-person possessive morphology and adjectival modification[2] of the noun. Data concerning positions requiring Dat, Instr or Comit case are less straightforward and we omit them.

(19) bʲi {nom-iːjə} / {honʲin nom} unʃ-aː
I book-ACC / interesting book read-PST
'I read a(n interestinxg) book' (#books)

(20) {nom-ʃni} / {ulaːn nom honʲin} baigaː
book-2SG / red book interesting was
'Your book / The red book was interesting' (#books)

We take these facts to mean that number interpretation is sensitive to the size of the DP structure the unmarked noun is part of – extended structure requires atomic semantics. We build our analysis on this suggestion.

5 Analysis

We propose that morphologically unmarked DPs in Buriat are structurally ambiguous. They can either lack the projection hosting number morphology or have a silent singular morpheme in it (ignoring linearisation): $[_{DP}...[_{\surd}$ nom]] 'book' vs. $[_{DP}...[_{NumP}\ \varnothing\ [_{\surd}$ nom]]] 'book-SG' (we locate NumP below DP following Farkas and de Swart 2010 a.o.). Thus we conjecture that in Buriat, the lack of NumP layer does not preclude the formation of DP (unlike, maybe, in some other languages). All DPs with overt plural morphology contain a NumP layer: $[_{DP}...[_{NumP}$ uːd $[_{\surd}$ nom]]] 'book-PL'. The interpretations of the three relevant substructures are the following: the form without the number projection has number-neutral interpretation:

[2]There is a certain amount of inter-speaker variation in whether adjectival modification precludes number-neutrality. Within our system, it may signal different attachment sites of adjectives in individual grammars.

(21) a. $[\![[_{\surd}\ \text{nom}]]\!] = \lambda x.{}^*\textsc{book}(x)$
b. $[\![[_{NumP}\ \varnothing\ [_{\surd}\ \text{nom}]]]\!] = \lambda x.\textsc{book}(x)$
c. $[\![[_{NumP}\ \text{u:d}\ [_{\surd}\ \text{nom}]]]\!] = \lambda x.{}^*\textsc{book}(x)$

The argument for such solution comes from data in (19) and (20) that suggest that extended syntactic structure correlates with the strictly atomic reading of unmarked forms. We think that this has to do with syntactic requirements of certain elements of DP structures. Namely, we suggest that adjectives, possessive morphology and certain case markers can't merge in the absence of NumP. We treat this fact as strictly syntactic.

The marked/unmarked direct object contrast buttresses this argument. Although in general the DO position can remain unmarked for Acc case (the conditions under which it happens are orthogonal to our point), whenever Acc is present, nouns with non-atomic reference have to host a plural morpheme and nouns not overtly marked for number denote strictly within atoms. This suggests that Acc selects for NumP. Similarly, (19)-(20) suggest that adjectival modification and possessive morphology generally require NumP to be projected.

Given the structures and meanings in (21), we can ask which of them can and do serve as alternatives to each other. $[_{\surd}$ nom] is strictly the least complex of these structures (the other two can be transformed into it by deletion of the Num head) – therefore, $A_{str}(...[_{\surd}$ nom]...) will be empty. $A_{str}(...[_{NumP}\ \varnothing\ [_{\surd}$ nom]]...) and $A_{str}(...[_{NumP}$ u:d $[_{\surd}$ nom]]...) can in principle contain the other two forms, as they are at most as complex – either of the same complexity, or, in the case of $[_{\surd}$ nom], strictly less complex. However, regardless of the entailment properties of the environment, $[_{\surd}$ nom] can't be kept as an alternative to *nom-u:d* 'book-PL': as they are synonymous, the negation of the sentence containing $[_{\surd}$ nom] would contradict the original sentence. There is no such dependence in (22-b):

(22) a. $A_{str}(...[_{NumP}$ u:d $[_{\surd}$ nom]]...) = { [...$[_{NumP}\ \varnothing\ [_{\surd}$ nom]] ...]}
b. $A_{str}(...[_{NumP}\ \varnothing\ [_{\surd}$ nom]]...) = { [...$[_{\surd}$ nom] ...], [...$[_{NumP}$ u:d $[_{\surd}$ nom]] ...]}

In sum, the plural (${}^*P$) form invokes the singular (P) form as an alternative; the singular (P) form invokes ${}^*P$ forms as alternatives. In this way, the problem of non-singular inferences of plural DPs via competition of two ${}^*P$-denoting forms does not arise – these forms are not in competition. This system is not very different from English, except for the existence of one more ${}^*P$ form as an alternative to the singular one (22-b). This unmarked alternative is not always active: sometimes, using the unmarked form instead of the singular one will result in ill-formedness due to requirements of other elements in the DP (case or possessive morphology or the adjective), but sometimes not. Even in the latter case, the semantic relationship between the source and its alternatives is never ${}^*P$ vs. ${}^*P$.

We do not argue here for any particular flavour of a strong sg / weak pl theory deriving non-singularity inferences of plurals by some pragmatic mechanism – be it scalar implicature (Spector, 2007; Zweig, 2009) or Maximize Presupposition (Sauerland, 2003; Sauerland et al., 2005). Rather, we point out that the underlying properties of the number system in Buriat, although it looks quite exotic, turn out surprisingly similar to that in English, and is reducible to it with the help of structurally filtered alternatives.

6 Extensions and discussion

We discussed Buriat data that seemed quite puzzling on the face of it – both number forms have number-neutral interpretation, but plural forms also show non-singularity inferences in UE

contexts. WEAK SG / STRONG PL theories have a problem covering Buriat data, STRONG SG / WEAK PL theories in combination with structural constraints on alternatives look promising.

Extensions of the analysis should cover 1) animate nouns (we talked about inanimate nouns only so far); 2) nominal number in quantificational DPs (with numerals, *many*, *all* etc.).

Animate nouns without plural morphology in Buriat range strictly over atoms. We encode this as a lexical requirement of animate nouns to project NumP.

Independent evidence for the presence of NumPs always projected by animate nouns, but not necessarily by inanimate nouns, comes from the distribution of agreeing demonstrative pronouns. Buriat demonstratives have distinct plural and singular (unmarked) forms: "ədə" 'these' and "ənə" 'this'. With animate nouns, it looks like demonstratives agree with the NumP, but the plural agreement is optional (verbal number agreement in Buriat is also optional), and the unmarked form "ənə" could be used on a par with the plural "ədə":

(23) bʲi ədə/ənə xʉbʉ:-d-i:jə xar-a:-b
I this.PL/SG boy-PL-ACC see-PST-1SG
'I saw these boys'

The pattern with inanimate nouns is more intricate. Most interestingly, the plural form "ədə" can appear with inanimate singulars, leading to the plural interpretation.

(24) basaga:-d xurgu:li-da: ədə nom asar-a:
girl-PL school-DAT.REFL this.PL book bring-PST
'The girls brought these books to their school'

The plural form "ədə" is incompatible with morphologically unmarked animate nouns.

(25) *bʲi ədə xʉbʉ: xar-a:-b
I this.PL boy.ACC see-PST-1SG
Intended: 'I saw these boys'

The contrast could indicate that with the inanimate nouns, "ədə" does not manifest agreement, but rather spells out the Num head (valued as "plural"). This option is not available for the animates, since they always project NumP independently, with demonstratives in a different syntactic position.

If we hypothesize further that the agreeing demonstratives only combine with phrases that have a NumP layer, we predict that we will not find the singular form "ənə" with unmarked inanimate nouns with non-atomic reference. This prediction is borne out:

(26) basaga:-d xurgu:li-da: ənə nom asar-a:
girl-PL school-DAT.REFL this.SG book bring-PST
'The girls brought this book/ #these books to their school'

In principle, "ənə" could either manifest agreement with the NumP or spell-out the Num head itself. But since, as we argue, unmarked number-neutral inanimates lack the NumP, only the latter option is available for them. Thus, in (26) the presence of "ənə" clearly signals that the value of Num is "singular", which is incompatible with the number-neutral interpretation.

Finally, we won't have much to say about the combinations of nouns with numerals and nominal quantifiers. Numerals combine with all three number options: unmarked, SG (secured by overt ACC marking) and PL:

(27) seren gurban nom / nom-u:d-i:jə / nom-i:jə-mni unʃ-a:
Seren three book / book-PL-ACC / book-ACC-1SG.POSS read-PST

'Seren read three books / my three books.'

We suggest that selectional restrictions of nominal quantifiers don't necessarily have consequences for the semantics of nominal number.

References

Bale, A. and H. Khanjian (2014). Syntactic complexity and competition: The singular-plural distinction in Western Armenian. *Linguistic Inquiry 45*(1), 1–26.

Bale, A., H. Khanjian, and M. Gagnon (2010). Cross-linguistic representations of numerals and number marking. In *Proceedings of SALT 20*, pp. 1–15.

Dayal, V. (2011). Hindi pseudo-incorporation. *Natural language and linguistic theory 29*(1), 123–167.

Farkas, D. and H. de Swart (2003). *The semantics of incorporation: from argument structure to discourse transparency*. Stanford, CA: CSLI.

Farkas, D. and H. de Swart (2010). The semantics and pragmatics of plurals. *Semantics and Pragmatics 3*, 1–54.

Fox, D. and R. Katzir (2011). On the characterization of alternatives. *Natural Language Semantics 19*(1), 87–107.

Katzir, R. (2007). Structurally-defined alternatives. *Linguistics and Philosophy 30*, 669–690.

Landman, F. (1991). *Structures for semantics*. Kluwer.

Landman, F. (2000). *Events and plurality*. Kluwer.

Link, G. (1983). The logical analysis of plurals and mass terms: A latticetheoretical approach. In R. Bauerle, C. Schwarze, and A. von Stechov (Eds.), *Meaning, use and interpretation of language*. Berlin: De Gruyter.

Mithun, M. (2010). *Constraints on compounding and incorporation*, pp. 37–56. Amsterdam: John Benjamins.

Öztürk, B. (2005). *Case, referentiality and phrase structure*. John Benjamins Publishing Company.

Sauerland, U. (2003). A new semantics for number. In R. B. Yound and Y. Zhou (Eds.), *Proceedings of SALT 13*, Ithaca, NY, pp. 258–275. Cornell Linguistics Club.

Sauerland, U., J. Andersen, and K. Yatsushiro (2005). The plural is semantically unmarked. In S. Kepser and M. Reis (Eds.), *Linguistic evidence*, pp. 413–434. Mouton de Gruyter.

Spector, B. (2007). Aspects of the pragmatics of plural morphology: on higher-order implicatures. In U. Sauerland and P. Stateva (Eds.), *Presuppositions and implicatures in compositional semantics*, pp. 243–281. New York: Palgrave-Macmillan.

Sudo, Y. (2017). Another problem for alternative-based theories of plurality inferences: the case of reduplicated plural nouns in japanese. *Snippets 31*, 26–28.

Zweig, E. (2009). Number-neutral bare plurals and the multiplicity implicature. *Linguistics and Philosophy 32*, 353–407.

Distributive numerals in Basque*

Patricia Cabredo Hofherr[1] and Urtzi Etxeberria[2]

[1] UMR 7023 - Structures formelles du langage
U. Paris Lumières /U. Paris 8 / CNRS, Paris, France
pcabredo@univ-paris8.fr
[2] CNRS - IKER, Bayonne, France
u.etxeberria@iker.cnrs.fr

Abstract

This paper presents the first detailed study of distributive numerals in Basque. We show that Basque distributive numerals are subject to restrictions on the obligatory licensing plurality that are not attested for distributive numerals in other languages described in the literature. We analyse the Basque NPs headed by distributive numerals as syntactically deficient noun-phrases that are semantically incorporated with an added requirement that the event be an event plurality satisfying a particular cumulation condition. We analyse the non-rigidity condition on Basque distributive numerals as an ignorance/ indifference condition on the referent, not as a plurality condition on the referents of the distributive numeral.

1 Introduction

Distributive numerals are a subclass of dependent indefinites introduced by lexically marked numerals. Dependent indefinites are defined as indefinites that impose a condition that their reference be non-rigid [Farkas, 1997, sect.4]. Here we present the first detailed study of distributive numeral NPs in Basque, marked by the suffix *-na* on the numeral (***num-na NPs***).[1]

Distributive numerals from a range of typologically diverse languages have been the object of a number of studies in the recent literature (Georgian [Gil, 1988], Hungarian [Farkas, 1997, Farkas, 2015], Romanian [Farkas, 2002], Telugu [Balusu, 2006], Kaqchikel Maya [Henderson, 2014], Tlingit [Cable, 2014], Serbocroatian [Knežević, 2015], ASL [Kuhn, 2015]). We show that Basque num-na NPs differ from the distributive numerals described in the literature with respect to their licensing profile. We analyse Basque num-na NPs as syntactically deficient noun-phrases that are semantically incorporated. We further argue that the non-rigidity condition on Basque num-na NPs is an ignorance/ indifference condition on the referent, similar to the identity of the implicit agent in *The chair was lifted twice*, not a plurality condition on the distributed share, as in proposed in the analyses for other languages.

We will proceed as follows. Section 2 outlines the syntactic distribution of num-na NPs. Section 3 presents the licensing conditions for Basque num-na NPs contrasting them with other distributive numerals described in the literature. Section 4 develops the analysis.

*Acknowledgements: This work is part of the project *The expression of (co-)distributivity cross-linguistically* of the Fédération Typologie et universaux du langage (CNRS 2559) and we would like to thank the participants of the project seminar for comments and discussion of previous versions of this work. We further thank the audiences at the Workshop on CoDistributivity 2017, and the participants of research seminars at the U. Pompeu Fabra, Barcelona, the IKER-group in Bayonne for comments and suggestions. The work of U. Etxeberria is supported by the following grants: IT769-13 (Basque Government), EC FP7/SSH-2013-1 AThEME 613465 (European Commission), FFI2014-51878-P and FFI2014-52015-P (Spanish MINECO).

[1] See [Euskaltzaindia, 1993, Rijk, 2008, Etxeberria, 2012] for basic descriptions of the suffix -na.

2 The syntax of num-na NPs

The Basque distributive suffix *-na* combines with numerals[2] and the wh-word *zenbat* "how much" [Rijk, 2008, 850].[3] In what follows we focus on noun phrases containing numerals+na.

(1) Ikasleek irakasleari **zazpi-na lan** aurkeztu zizkioten.
student-D.pl.erg teacher-D.sg.dat seven-na work.abs present aux.pl
The students presented seven works each to the teacher. [Etxeberria, 2012, 55: ex 203]

(2) **Zenba**-na filma ikusi zituzten hiru umeek?
how.many-na film watch aux three child-D.pl.erg
How many films each did the three children watch?

2.1 The syntactic distribution of num-na NPs

Num-na NPs cannot be subjects (3) [Trask, 2003, 128]. [Rijk, 2008, 852] gives one attested example with the verb *help* and distribution over a 1pl dative experiencer (4). However, this example is not acceptable to our informants[4] and there are no examples of num-na NPs in subject position in the corpus **Mendeko Euskararen Corpus Estatistikoa**.[5]

(3) *__Bi-na umek__ hiru tarta jan zituzten.
two-na kid.erg three cake eat aux
Intended: The three cakes were eaten by two kids each.

(4) % **Bos-na gizonek** lagundu digute. (G.B.2 89)
five-na man-D.pl.erg help aux
Each of us (dative) was helped by five men. [Rijk, 2008, 852]

Num-na NPs can be direct (1) and indirect objects (5-a), PP complements with case-markers *-etan/-ekin* "locative/with" (5-b)/(5-c), PP adjuncts (5-d) and noun complements (5-e)/ (5-f).

(5) a. Ikasleek **zazpina irakasleri** lan bat aurkeztu zieten.
student.D.pl.erg seven-na teacher.dat work one present aux (indirect obj.)
The students presented one work to seven teachers each. [Etxeberria, 2012, 55]

b. Jonek neskei [**bi-na igandetan**] lan egin arazi zien.
Jon-erg girl-D.pl.dat **two-na Sunday-in** work do cause aux
John made each of the girls work on two Sundays. (locative case)
Lit. John made the girls work on two Sundays each.

c. Mutilek **hiru-na enbaxadorerekin** hitzegin zuten.
boy-Dpl.erg three-na ambassador-with talk aux
The boys spoke with three ambassadors each. (PP complement)

d. Bi liburu oso garesti erosi zituen **bi-na lagunekin**.
two book very expensive buy aux two-na friend-D.pl-with
He bought two very expensive books with two friends each. (PP adjunct)

[2]The distributive numeral preserves the syntactic position of the simple numeral: *bat/ba-na* "one/one-na" is post-nominal while *bi/bi-na "two/two-na"* and other numerals are pre-nominal [Etxeberria, 2012].

[3]De Rijk also includes fractionals. However, as in fractionals *-na* combines with the numeral denominator: ***erdi ba**-na* "half one-na", and not with the fractional itself **erdi-na* "half-na" we leave them aside here.

[4]% marks dialectal variation in acceptability. See [Knežević, 2015] who observes that in SerboCroatian only some speakers accept distributive numeral *po* phrases in subject position

[5]Corpus **Mendeko Euskararen Corpus Estatistikoa** http://xxmendea.euskaltzaindia.eus/Corpus/

e. Mutilek **hiru-na herrialdetako** enbaxadorerekin hitzegin zuten.
boy-Dpl.erg three-na country-gen ambassador-with talk aux
The boys spoke with the ambassadors of three countries each. (complement of N)

f. Umeek **bi-na musikariren inguruko** liburu bat irakurri zuten.
child-D.pl.erg two-na musician.gen about book one read aux
The children each read one book about two musicians. (PP complement of N)

Also note that the noun introduced by numeral+na can be modified (6).

(6) Emakume-ak **bi-na** edari **hotz** / **limoiarekin** eman zizkien.
woman-D.erg two-NA drink cold / with-lemon give aux
The woman gave them two cold drinks/ two drinks with lemon each.

2.2 Distributive properties of num-na NPs

Following [Choe, 1987, 90] we distinguish the SORTAL KEY and the DISTRIBUTED SHARE. Choe analyses distribution as a quantificational relationship between the atoms of the key and the share. In Choe's terms distributive numerals mark the **distributive share** (share markers).

(7) a. Distributive dependency is a relation between a **sortal key** A and **distributed share** DistShare B. [Choe, 1987, 90].

b. $\forall$ a $\in$ **A** : atom(a) $\exists$ **B** : R(a,B)
Key DistrShare

Num-na NPs allow (6) but do not require temporal distributivity: they combine with stative predicates (8) and are acceptable in contexts without temporal distribution (9-a). Num-na NPs allow overlapping distributivity (9-b) (example adapted from [Knežević, 2015]) but exclude readings in which the sortal key provided by the licensing plurality is not exhausted (9-c).

(8) Neskek **bi-na hizkuntza** dakizkite
girl-Dpl.erg two-na language know
The girls know two languages each.

(9) Umeek **bina globo** eutsi zituzten.
child-D.pl two-na balloon hold aux.past
The children held 2 balloons each.

a. ok: Context 1: children each holding 2 balloons (no temporal distribution)
b. ok: Context 2: children holding two balloons, with two children holding the same balloon (plus another balloon in the other hand).
c. excluded: Context 3: one of the children is not holding any balloons

2.3 Relative scope of distributive numerals

As dependent indefinites, distributive numerals depend on a plurality for their interpretation and take narrow scope with respect to their licensing plurality [Farkas, 2015]. Num-na NPs only take narrow scope with respect to modals (10-a), negation (11-a) and quantifiers (12-a).

(10) a. Umeek **bina poema** irakurri **behar** dituzte.
child.pl.erg two-na poem read must aux
ok narrow scope: The children have to read two poems each.
*intermediate scope: For each child there are two poems that the child must read.

b. Umeek **bi** **poema** irakurri **behar** dituzte.
child.pl.erg two poems read must aux
ok narrow scope / ok intermediate scope

(11) a. Umeek **ez** zituzten **bina** **liburu** irakurri.
kid.D.pl.erg neg aux two-na book read.
The children did not read two books each.
Ok Narrowest scope: It is not true that the children read two books each.
*Wide scope wrt neg + narrow scope wrt the children: There are two poems for each child that s/he did not read.

b. Umeek **ez** zituzten **bi** **liburu** irakurri.
kid.D.pl.erg neg aux two book read
Ok Narrowest scope / Ok Wide scope wrt neg + narrow scope wrt the children

(12) a. Jonek eta Mirenek **bi-na** **pasahitz** jartzen dituzte beraien email kontu
Jon.erg and Miren.erg two-na password put.prog aux their email account
bakoitzean.
every-D.sg-in
ok narrow scope: Jon and Miren put two passwords into every email account.
* intermediate scope: Jon and Miren have two passwords each that they put into every email account.

b. Jonek eta Mirenek **bi** **pasahitz** jartzen dituzte beraien email kontu
Jon.erg and Miren.erg two password put.prog aux their email account
bakoitzean.
each-D.sg-in
ok narrow scope / ok intermediate scope

Unlike num-na NPs, simple indefinites allow intermediate readings (10-b)/ (11-b)/ (12-b).

3 Licensing distributive numerals in Basque

Distributive numerals depend on a plurality for their interpretation. However, the types of plurality that can fulfill the licensing requirements for distributive numerals vary crosslinguistically ([Farkas, 1997, Balusu, 2006, Henderson, 2014, Cable, 2014, Knežević, 2015, Kuhn, 2015]).

Here we show that the licensing conditions for num-na NPs are different from the licensing profiles for other distributive numerals described in the literature.

Num-na NPs are ungrammatical without an overt licensor (13-a), like binominal *each* and Kaqchikel distr-num [Henderson, 2014], and contrasting with distributive numerals in Telugu (13-b), Tlingit and SerboCroatian [Balusu, 2006, Cable, 2014, Knežević, 2015], that can be licensed by implicit distribution over times/ locations.

(13) a. *Ne-re seme-ak **hiru-na** arrain harrapatu zituen. (Basque)
I-gen son-D.erg three-dist fish catch aux.past
Intended: My son caught three fishes (each time/ on each occasion).

b. Raamu **renDu** **renDu** kootu-lu-ni cuus-ee-Du (Telugu)
Ram 2 2 monkey-Pl-Acc see-Past-3PSg
a. Ram saw 2 monkeys (in each time interval). Implicit temporal key
b. Ram saw 2 monkeys (in each location). Implicit spatial key [Balusu, 2006, ex9]

Num-na NPs are licensed by plural and quantified co-arguments (14-a). The complements have

to be clause-mates: licensing into an embedded predicate across a perception verb is impossible (14-b). Num-na NPs are also licensed by plural locative adjuncts (15).

(14) a. Ume guztiek / Umeek **bina** liburu irakurri zituzten.
child all.D.pl.erg / children.erg two-na books read aux
All the children / The children read two books each.
b. *__Mutil-ek__ [Maria **bi-na** pizza erosten] ikusi zuten.
boy-erg.pl Maria two-na pizza buy.prog see aux
Intended: The boys each saw Maria buy 2 pizzas.

(15) Jonek bi-na liburu erosi ditu liburudenda guztietan / horietan.
Jon.erg two-dist book buy aux bookstore **all-D.pl-loc** / **those.pl-loc**
Jon bought two books in each of **all the** / **those bookstores**. (apud [Rijk, 2008, 852])

In contrast with other languages that allow licensing by adverbial expressions, adverbs like *beti* "always" and when-clauses do not license num-na NPs (16)/(17) (≠ Kaqchikel dist-num allowing *always*). More precisely, **unbounded** temporal adjuncts do not license num-na NPs (18-a): the adjunct has to be bounded (18-b).

(16) *Manuelek **beti** bina pizza jaten ditu
Manuel-erg **always** two-na pizza eat-hab aux
Intended: Always/on each occasion M. eats 2 pizzas.

(17) *Ni ikustera etortzen denean Manuelek bina opari ekartzen dizkit
me-abs see-nmz-all come-hab is-rel-loc Manuel-erg two-na present bring-hab aux
Intended: **When he comes to see me**, Manuel brings me two presents.

(18) a. ***Igandetan**, Manuelek bi-na opari ekartzen zizkidan.
Sunday-loc.pl Manuel.erg 2-na present bring aux (to me)
Intended: On Sundays M. brought/used to bring me two-na presents.
b. **Azken bi igandeetan** Manuelek bi-na opari ekarri dizkit.
last two Sunday-loc.pl Manuel.erg 2-na present bring aux (to me)
The last two Sundays M. brought me 2 presents each time.

In the parallel examples with an **unmarked** indefinite, a dependent reading wrt to the unbounded temporal plurality is possible (19), showing that the num-na NPs in (16)/(17)/(18-a) are in the semantic scope of the unbounded temporal expressions.

(19) a. Manuelek **beti bi** pizza jaten ditu
Manuel-erg **always two** pizza eat-hab aux
Manuel always eats two pizzas. i.e. on each relevant occasion M. eats 2 pizzas.
b. Ni ikustera etortzen denean Manuelek **bi** opari ekartzen dizkit
me-abs see-nmz-all come-hab is-rel-loc Manuel-erg two present bring-hab aux
When he comes to see me, Manuel brings me two presents.
c. **Igandetan**, Manuelek **bi** opari ekarri dizkit.
Sunday-loc.pl Manuel.erg two present bring aux (to me)

The licensing plurality can be in an argumental PP (20-a) but not in an adjunct PP (20-b).

(20) a. Jonek **espezialista hauekin** bi-na arazoz hitzegin zuen.
Jon.erg specialist these-with two-na problem-instr talk aux.
Jon spoke about two problems with each of these specialists . (argument PP)

b. *?Jonek bina liburu erosi zituen **umeekin**.
Jon.erg two-na book buy aux children-with (adjunct PP)
Intended: Jon bought two books with each of the children. (Lit. Jon bought two-na books with the children).

[Farkas, 2015] points out that pluralities of worlds do not license distributive numerals. This also holds for Basque: generics (21-a) and modals (21-b) do not license num-na NPs.

(21) a. ***Txakurrek** lau-na hanka dituzte.
dog-D.pl four-na leg have
Not: Dogs have four legs. (generic subject), ok with anaphoric definite *the dogs*
b. *Mirenek bina liburu irakurri **behar** ditu.
Miren.erg two-na book read must aux
Not: Mari must read two books. (modals)

4 Analysis

Our analysis of num-na NPs involves three elements: (i) num-na NPs are semantically incorporated ([Chung & Ladusaw, 2004]), (ii) num-na NPs mark the event predicate they combine with as pluractional and (iii) the event plurality is subject to a restriction requiring it to be the cumulation of a bounded sum of the sub-events that are indexed by the plural licensor.

Following [Chung & Ladusaw, 2004] we assume that there are two modes of composition for a noun-phrase: Restrict and Saturate. We analyse num-na NPs as syntactically deficient noun-phrases that are semantically composed via Restrict [Chung & Ladusaw, 2004].[6] The semantically incorporated num-na NP is interpreted as a predicate modifier introducing a sortal restriction by the N and a cardinality restriction by the numeral bearing on the theta-role corresponding to the argument position occupied by the NP. The num-na NP does not introduce a discourse referent: the argument position is bound off by an existential closure operation EC ([Chung & Ladusaw, 2004]).

(22) a. RESTRICT applied to a two place predicate [Chung & Ladusaw, 2004, 10]
Restrict $((\lambda\ x\ \lambda\ y\ [Verb(x,y)], P(y)) = (\lambda\ x\ \lambda\ y\ [Verb(x,y)\ \&\ P(y)]$
b. Existential Closure $(\lambda\ x\ \lambda\ y\ [P(x,y)]) = \lambda\ x\ \exists\ y\ [P(x,y)]$

We further propose that num-na NPs function as event modifiers that contribute a dependency between the plural licensor and the event-description containing the num-na NPs, with an additional requirement that the event plurality be a bounded sum of events co-indexed with a plural licensor. The co-indexing plurality has to be an argument or a locative or temporal adjunct of the event description containing the num-na NP.

(23) a. Umeek **bina** aulki altxatu dituzte.
child-D.pl two-na chair lift aux
The children lifted two chairs each.
b. $\exists\ E \in lift^*(e,x,y)$: $E = \Sigma\ (e_i$: $x_i \in$ [[children]] & $\exists y\ P(e_i,x_i,y)$ & two-chairs(y))
c. There is a event-plurality E such that E is the sum of individual events with the num-na NP semantically incorporated into its argument position and existentially bound
and the event plurality can be indexed by the atoms of the plural licensor DP.

[6] The num-na NPs are clearly not morphologically incorporated as the noun can be modified cf. (6).

(24) a. Jonek bi-na liburu erosi ditu liburudenda horietan.
Jon.erg two-dist book buy aux bookstore **those.pl-loc**
Jon bought two books each in those bookstores.
b. $\exists$ E $\in$ buy*(e,j,y,l): E = Σ (e_i: l_i $\in$ [[those bookstores]] & $\exists$y buy(e_i,j,y,l_i) & two-books(y))

The semantics proposed here is modification of the event description combined with an explicit cumulation condition on the event plurality indexed to the licensor. This is akin to a lexical expression of distributivity similar to that of a NP modified with the adverb *respectivamente* "respectively" in the following example from Spanish.

(25) 9 han traducido dos obras cada uno y 3 traductores han traducido
9 translated two works each one and 3 translators have translated
otros tres libros respectivamente (Spanish)
another three books respectively http://www.academia.edu/26224464/

(26) a. Los niños leyeron dos libros respectivamente.
The children read two books respectively.
b. Juan y Ana hablaron con los embajadores de tres países respectivamente.
John and Ana spoke with the ambassadors of three countries respectively.

Distributive configurations are pairings of the atoms of the sortal key with elements corresponding to the description of the share. Quantification by universal quantification over an existential quantification in the syntax is only one way of achieving such a pairing.[7] Other ways of imposing a paired structure between the sortal key and share are world knowledge (27-a) and lexical modification (27-b). Our analysis places the distribution contributed by num-na NPs on the side of lexical modification.

(27) a. The children arrived on tandems. $\longrightarrow$ in groups of two. (World knowledge)
b. The children arrived in pairs. $\longrightarrow$ in groups of two. (Lexical modification)

Semantic incorporation accounts for the fact that num-na NPs have narrowest scope (section 2.3) and cannot be taken up in the following discourse (28).

(28) Azken bi igandeetan Manuelek bina liburu ekarri dizkit.
last two sunday-loc.pl Manuel.erg 2-na present bring aux.
The last two Sundays M. brought me 2 books each time.
a. #Apal horretan gorde ditut.
shelf that.loc hide aux.3plA-1sE
I put them on that shelf. (*them* = null argument + agreement on aux)
b. #Nere izena idatzi dut **beraien** barnean.
my name write aux they-loc inside-loc
I wrote my name inside of them. (possessive pronoun)

The ban on num-na NPs in subject position (3) is also characteristic for many other instances of semantically incorporated arguments.

The fact that num-na NPs can appear in non-additive measure-phrases (29) further confirms that num-na NPs do not introduce a referent. However, based on [Laca, 1990] observing that psych-predicates only take individuals as arguments, we would expect that semantically incorporated noun phrases like num-na NPs are not possible with these psych-predicates. This

[7] We thank Hans Kamp for pointing this out to us.

is only partially borne out however as the examples in (30) are not completely ungrammatical. Speakers seem to have shifting grammaticality judgements with these sentences. Two speakers accepted num-na NPs in an anchored context, i.e. (30-b) said coming back with the children from the zoo, but found it degraded as a general preference statement with the imperfective verb form (30-c).[8] We have no explanation for this contrast.

(29) Ontziak **36na gradutan** egon behar dute laborategi honetan
recipient-D.pl.erg 36-na degree.loc be must aux laboratory this.loc
The recipients have to be at 36 degrees each in this laboratory.

(30) a. ?Neskek **bina barazki** gorroto dituzte.
girl-D.pl two-na vegetable hate aux
The girls dislike two vegetables each.
b. Nere semeei **bina animalia** gustatu zaizkie.
I-gen son-D.pl-dat animal like aux
My sons liked two animals each. (Said coming back from the zoo)
c. ?Nere semeei **bina animalia** gustatzen zaizkie.
I-gen son-D.pl-dat animal like-hab aux
My sons like two animals each.

A simple numeral and a num-na NP can be coordinated (31-a); this is not problematic for an incorporation account as a full DP and a semantically deficient NP can be coordinated (31-b).

(31) a. Ikasleek patata tortila haundi bat eta sardeska **bana** eskatu zituzten.
student-erg potato omelette big one and fork one-na ask aux
The students asked for one big omelette (for everyone) and one fork each.
b. Gaur goizean egunkaria eta zure gutuna irakurri ditut.
today morning-in newspaper and your letter read aux
This morning I read the newspaper and your letter.

The locality conditions on the dependency introduced by num-na NPs are partly similar to the locality of the antecedent for internal readings of *desberdin bat* "different.sg one", that also allows plural temporal and locative adjuncts (32)/ (33) but not licensing by coordinated verbs (34). However, the locality conditions on the licensor for num-na NPs and for internal readings of *desberdin bat* differs in examples with an embedding perception verb (35).

(32) Jonek filma **desberdin bat** ikusi zuen astelehenean eta asteartean.
Jon.erg film different one see aux Monday.in and Tuesday.in
Jon saw a different film on Monday and on Tuesday. cf. (18-b)

(33) Jonek liburu **desberdin bat** erosi ditu liburudenda guztietan.
Jon.erg book different one buy aux bookstore **all-D.pl-loc**
Jon bought a different book in each of **all the** / **those bookstores**. cf. (15)

(34) a. *Jonek filma **desberdin bat** ikusi eta kritikatu zuen.
Jon.erg film different one see and criticise aux
b. *Jonek **bi-na** filma ikusi eta kritikatu zituen.
Jon.erg two-na films see and criticise aux

(35) a. Mutil guztiek Miren soineko **desberdin bat** erosten ikusi zuten.
boy all-D-pl.erg Miren dress **different one** buying see aux

[8] This may be related to the fact that the examples of num-na NPs with stative predicates are all with s-level statives *be at 36 degrees/ hold two balloons*. The difference (30-b) vs. (30-c) is an s-level/i-level contrast.

Each of the boys saw Mary buy a dress and the dresses were different.

b. *Mutil guztiek Miren **bina** soineko erosten ikusi zuten.
boy all-D-pl.erg Miren two-na dress buying see aux

Basque num-na NPs are possible with non-additive degree expressions (29), showing that there is no plurality requirement on referents corresponding to the num-na NPs. Also, in cases where the speaker is ignorant about the identity of the instantiations of the share, num-na NPs are felicitous, even if the referents corresponding to it in the scenario happen not to be different.

(36) Umeek **bina** aulki altxatu dituzte.
child-D.pl two-na chair lift aux
The children two-na chairs.

a. Context 1: Different photos, each depicting one child lifting two chairs.
⟶can use (36) even if the chairs happen to be the same two chairs for everyone.

b. Context 2: One scene with the children taking turns lifting the same two chairs.
⟶cannot use (36)

The identity of the instantiations of the argument corresponding to num-na NP is **unspecified**, not **specified as varying**. This is comparable to the implicit agent in *The door was opened twice* or the implicit themes of *John read and Mary read* where it is left unspecified whether the agents/ books read are different or not. We therefore do not adopt a plurality presupposition/ postsupposition on the num-na NP as proposed in Balusu's 2006, Henderson's 2014 or Farkas's 2015 analyses of distributive numerals. However, explicit knowledge of the identity of the referent of the share (e.g. by direct visual evidence of the scenario) seems degraded, so there seems to be an ignorance condition attached to the identity of the share.

The situation seems to be similar for English binominal *each*. In ex. (37), it is not *necessary* that there be more than two films watched, (37) asserts that there is an **event plurality** composed of subevents of *watching two films* for each child. The identity of the referents corresponding to *two films* can be blurred across the event-plurality as in event-related readings of examples like *4000 ships passed through the lock* [Krifka, 1990].

(37) The children watched two films each.

5 Conclusion

According to the analysis proposed here num-na NPs are syntactically deficient noun-phrases that are interpreted by semantic incorporation. As such they contribute a predicate modification to the argument position they occupy without introducing a discourse referent corresponding to the num-na NP. Num-na NPs impose two further conditions: (i) a dependency condition between a plural licensor and the event-description that contains the num-na NP and (ii) an ignorance condition wrt to the identity of the indefinite. This analysis follows [Farkas & de Swart, 2004] and [Chung & Ladusaw, 2004] in distinguishing the lexical use of variables (as argument positions) and the discoursive use of variables as discourse referents.

Analyses of distributive numerals in other languages treat them as introducing a discourse referent. The analyses proposed for Telugu [Balusu, 2006] and Tlingit [Cable, 2014] furthermore include a distributive component introduced by the distributives numerals themselves; this does not carry over to num-na NPs, auto-licensing by an implicit plurality is impossible (13-a).

The proposal in [Farkas, 2015] treats the variable introduced by the distributive numeral NP as dependent on a licensing variable. This account does not carry over to num-na NPs

since the dependent variable account derives narrow scope of the distributive numeral NP **wrt the licensor**, but not with respect to other scope taking elements and therefore, intermediate scope readings are expected, running counter the behaviour of num-na NPs (10)/(11)/(12).

The analysis in [Henderson, 2014] relies on a condition that imposes that the variable of the distributive numeral NP is marked for evaluation plurality: the variable is interpreted by a set of assignment functions such that the value assigned to the NP containing the distributive numeral is not constant across the set of assignment functions. As we have shown, the non-rigidity condition in Basque is not a plurality condition on the domain of the distributive share but rather a condition that identity of the distributive shares must not be part of the context. A plurality condition imposes a condition that requires the existence of at least two different instances; the condition imposed for Basque num-na NPs is weaker than presupposed plurality: identity of the instantiations of the share is not excluded as long as the identity is not part of the context. The analysis of ignorance/indifferent conditions is also central in the analysis of epistemic indefinites like Sp. *algún +N* "some N or other" [Alonso-Ovalle & Menéndez-Benito, 2013], and num-na NPs should be examined in comparison with these types of indefinites.

References

[Alonso-Ovalle & Menéndez-Benito, 2013] Epistemic indefinites: Are we ignorant about ignorance? *Proceedings of the 19th Amsterdam Colloquium*.

[Balusu, 2006] Distributive reduplication in Telugu. In *Proc. of NELS 36*, 39–53.

[Cable, 2014] Distributive numerals and distance distributivity in Tlingit (and beyond). *Language*, 90(4):562–606.

[Choe, 1987] Choe, Jae-Woong. *Anti-quantifiers and a theory of distributivity*. PhD, UMass Amherst.

[Chung & Ladusaw, 2004] *Restriction and saturation*. MIT Press, Cambridge MA.

[Etxeberria, 2012] Quantification in Basque. In Keenan, E. and Paperno, D., editors, *Handbook of Quantifiers in Natural Languages*. Springer.

[Euskaltzaindia, 1993] *Euskal Gramatika Laburra: Perpaus Bakuna*. Euskaltzaindia [Academy of the Basque Language].

[Farkas, 1997] Dependent indefinites. In Corblin, F., Godard, D., and Marandin, J.-M., editors, *Empirical Issues in Syntax and Semantics 1*, 243–268. Peter Lang.

[Farkas, 2002] Extreme non-specificity in romanian. In Beyssade, C. et al., editors, *Romance Languages and Linguistic Theory 2000*, Amsterdam/ Philadelphia. John Benjamins.

[Farkas, 2015] Dependent indefinites revisited. Talk at *Journées (Co-)Distributivité 2015*, Paris.

[Farkas & de Swart, 2004] Incorporation, plurality and the incorporation of plurals: a dynamic approach. *Catalan J. of Linguistics*, 3:45–73.

[Gil, 1988] Georgian reduplication and the domain of distributivity. *Linguistics*, 26:1039–1065.

[Henderson, 2014] Dependent indefinites and their post-suppositions. *Semantics and Pragmatics*, 7.

[Knežević, 2015] *Numerals and Distributivity in Serbian: at the syntax-semantics-acquisition interface*. PhD, Université Nantes.

[Krifka, 1990] Four thousand ships passed through the lock: Object-induced measure functions on events. *Linguistics and Philosophy*, 13(5):487–520.

[Kuhn, 2015] *Cross-categorial and plural reference in sign language*. PhD, NYU.

[Laca, 1990] Generic objects : Some more pieces of the puzzle. *Lingua*, 81:25–46.

[Rijk, 2008] Rijk, R. P. d. *Standard Basque: A Progressive Grammar*. MIT Press.

[Trask, 2003] Trask, L. The noun phrase: nouns, determiners and modifiers; pronouns and names. In Hualde and Ortiz de Urbina, editors, *A Grammar of Basque*, 113–170. Mouton de Gruyter.

Homogenous Alternative Semantics

Fabrizio Cariani[1] and Simon Goldstein[2]

[1] Northwestern University
[2] Lingnan University

Abstract

We explore the interaction between conditional excluded middle and simplification of disjunctive antecedents. After showing these principles to be nearly incompatible, we develop an approach that fits in the narrow space they leave open.

1 Introduction

David Lewis's logic for the counterfactual conditional [19] famously invalidates two plausible-sounding principles: simplification of disjunctive antecedents (SDA),[1] and conditional excluded middle (CEM).[2] Simplification is the entailment: $(\mathsf{A} \text{ or } \mathsf{B}) > \mathsf{C} \vdash (\mathsf{A} > \mathsf{C}) \,\&\, (\mathsf{B} > \mathsf{C})$. For instance, given SDA, (1) entails (2).

(1) If Hiro or Ezra had come, we would have solved the puzzle.

(2) If Hiro had come, we would have solved the puzzle and if Ezra had come, we would have solved the puzzle.

As for CEM, it is the validity claim: $\vdash (\mathsf{A} > \mathsf{B}) \vee (\mathsf{A} > \neg\mathsf{B})$. A distinctive consequence of CEM is that the negation of $\mathsf{A} > \mathsf{C}$ entails $\mathsf{A} > \neg\mathsf{C}$. For instance, (3) entails (4).

(3) It is not the case that if Hiro had come we would have solved the puzzle.

(4) If Hiro had come, we would not have solved the puzzle.

Much attention has been devoted to these heretical principles in isolation, but relatively little work has considered their interaction. Since there are strong arguments for both principles, it is urgent to investigate how they might be made to fit.

Our pessimistic finding is that the heresies do not mix easily. We present a battery of incompatibility results showing that no traditional theory of conditionals or disjunction can allow them to coexist. Despite these negative findings, we argue that the project of combining CEM and SDA is not hopeless—provided that we are willing to incorporate insights from the linguistics literature within our framework for conditional logic. To validate both principles, we synthesize two tools that can be used to validate each principle individually: the alternative sensitive analysis of disjunction [1] and the theory of homogeneity presuppositions [11].

The resulting theory requires one last heresy: the entailment relation must be intransitive. In particular, while CEM is valid, other principles are invalid that are logical consequences of it.

2 The Case for the Heresies

The main argument for SDA seems to consist entirely in the observation that instances like the one from (1) to (2) sound extremely compelling (see [9, p.453-454]). Obviously, this is

[1] See [9], [10]; [21], [22]; [20].
[2] See [27], [11]; [30], and [16].

not a full defense of SDA, but it creates a strong presumption in its favor—one that would require substantial theoretical argument to be overthrown. Indeed, contemporary approaches in truth-maker semantics (e.g., [10]) are designed around the desire to validate it.

CEM is not typically justified by this direct method. Instead, its defenders propose that various phenomena fall into their proper place if we accept CEM's validity. For example, the inference from (3) to (4) turns out to be an application of disjunctive syllogism. More generally, conditionals with *will* and *would* consequents fail to enter into the scope relations that would be expected if CEM failed [27, p.137-139]. A recent version of this argument relies on data involving attitude verbs that lexicalize negation (see [5]).

(5) I doubt that if you had slept in, you would have passed.

(6) I believe that if you had slept in, you would have failed.

The equivalence is easily explained if CEM is valid (and assuming that failing equals not passing). The speaker doubts *sleep* > *pass*; if there was a way for this conditional to be false other than by *sleep* > *fail* being true, it should be possible to accept (5) without accepting (6). By contrast, it is hard, if not impossible, to explain without CEM. This argument streamlines an older argument for CEM involving the interaction between conditionals and quantifiers.[3] Consider:

(7) No student will succeed if he goofs off.

(8) Every student will fail if he goofs off.

(7) and (8) are intuitively equivalent. They appear to involve quantifiers taking scope over conditionals. Given CEM and this scope assumption, they are predicted equivalent. Take an arbitrary student, and suppose it is false of him that he will succeed if he goofs off. By CEM it follows that he will fail if he goofs off. On reflection, then, the interaction of conditionals and quantifiers also favors the validity of CEM.

Our final argument for CEM is based on the interaction between *if* and *only*.[4] CEM can help explain why *only if* conditionals imply their converses. Consider the following conditionals:

(9) The flag flies only if the Queen is home.

(10) If the flag flies, then the Queen is home.

(11) The flag flies if the Queen isn't home.

(9) entails (10). In [11] this entailment is derived compositionally, on the assumption that *only* in (9) takes wide scope to the conditional. *Only* then negates the alternatives to the conditional *the flag flies if the Queen is home*, which are assumed to include (11). Given some background assumptions, Conditional Excluded Middle and the negation of (11) imply (10).

3 Incompatibility Results

Having introduced our favorite conditional heresies, we show that they are in tension with each other. In keeping with a distinction we have drawn in the previous section, we appeal to two distinct notions of disjunction: (i) natural language *or* and (ii) Boolean disjunction, '$\vee$'. Given the asymmetry we highlighted in how SDA and CEM are justified, it will strengthen

[3] See [15]; [14]; [18]; and [16] for discussion.
[4] See [2] and [11] for discussion.

our argument to refrain from assuming that these have the same meaning. Our results require classical assumptions about the logic of '$\vee$' but very few assumptions about the meaning of *or*.

3.1 Collapse

CEM and SDA together imply collapse to the material conditional, given relatively modest logical assumptions. We assume standard sequent rules for classical connectives as well as the standard structural rules governing classical logic.[5] Among the structural rules, the transitivity of entailment will play a very important role in our discussion. Transitivity follows from Cut when X and Y are empty.

Cut. if $X \vdash \mathsf{A}$ and $Y, \mathsf{A} \vdash \mathsf{B}$, then $X, Y \vdash \mathsf{B}$

Several of our proofs rely on disjunction rules, so it is worth stating them explicitly

Cases. if $X, \mathsf{A} \vdash \mathsf{C}$ and $Y, \mathsf{B} \vdash \mathsf{C}$, then $X, Y, (\mathsf{A} \vee \mathsf{B}) \vdash \mathsf{C}$

$\vee$-Intro. if $X, \mathsf{A} \vdash \mathsf{B}$, $X, \mathsf{A} \vdash \mathsf{B} \vee \mathsf{C}$

To these, add specific assumptions about conditionals (three axioms and one rule). The axioms are *modus ponens* ($\mathsf{A}, \mathsf{A} > \mathsf{C} \vdash \mathsf{C}$), *reflexivity* ($\vdash \mathsf{A} > \mathsf{A}$) and *agglomeration* ($\mathsf{A} > \mathsf{B}, \mathsf{A} > \mathsf{C} \vdash \mathsf{A} > (\mathsf{B} \,\&\, \mathsf{C})$). As for the rule, it is:

Upper Monotonicity. if $\mathsf{B} \vdash \mathsf{C}$, then $\mathsf{A} > \mathsf{B} \vdash \mathsf{A} > \mathsf{C}$

While these assumptions are not entirely uncontroversial, they are generally accepted in the literature. For ease of reference, we call this combination of assumptions *the classical package*.

We can now state our result more precisely (Proofs of all results are omitted here. They are presented in [4]; Fact 1 is related, but not identical, to a result in [3]).

Fact 1. *Given the classical package,* CEM *and* SDA *imply that* $\mathsf{A} > \mathsf{C} \dashv\vdash \neg\mathsf{A} \vee C$.

Previous work on SDA has shown that it sits in major tension with the substitution of logical equivalents ([9], [8]). Interestingly, our own result makes no use of this principle. More generally, we assume nothing about the semantic or logical properties of *or*, except that it supports SDA.

3.2 Interconnectedness of all things

Our second result is that combining CEM and SDA forces the conditional to validate an undesirable schema, which we call IAT for "the Interconnectedness of All Things".

IAT. $(\mathsf{A} > \mathsf{C} \,\&\, \mathsf{B} > \mathsf{C}) \vee (\mathsf{A} > \neg\mathsf{C} \,\&\, \mathsf{B} > \neg\mathsf{C})$

Validating IAT is undesirable because it requires an extreme level of dependence among arbitrary distinct sentences. Suppose, for instance, that A="Abe flies", B="Bea runs" and C="Cleo swims". Then it must be that either both *Abe flies* $>$ *Cleo Swims* and *Bea runs* $>$ *Cleo swims* are true or both *Abe flies* $>$ *Cleo does not swim* and *Bea runs* $>$ *Cleo does not swim* are. Among other things, this appears to entail that it is incoherent to reject both of the following:

(12) If Abe flies, then Cleo swims.

(13) If Bea runs, then Cleo does not swim.

[5] For contemporary sources on the sort of system we presuppose, see [29] and [23].

It would be incorrect to say that no conditional validates IAT. For one thing, the material conditional does. Nonetheless, we comfortably assert that only unsatisfactory conditional connectives satisfy IAT. Here is an explicit statement of the second incompatibility result.

Fact 2. *Given disjunction rules, cut,* CEM*, and* SDA*,* IAT *must be a logical truth.*

3.3 *Might* conditionals

We end this section by noting a third result which, though slightly different in spirit, plays an important role in our theoretical discussion. Alonso-Ovalle [1] observes simplification with *might* conditionals (specifically counterfactuals), as in the inference from (14) to (15).

(14) If Hiro or Ezra had come, we might have solved the puzzle.

(15) If Hiro had come, we might have solved the puzzle.

Additionally, he shows that strict accounts of counterfactuals cannot validate this form of simplification, given a Boolean semantics for disjunction.

It will be convenient for our purposes to take *If A, might B* as idiomatic. Formally, we write this as $\mathsf{A} >_\diamond \mathsf{B}$. With this symbol in hand we state:

$\diamond$-SDA. $(\mathsf{A}\text{ or }\mathsf{B}) >_\diamond \mathsf{C} \vdash (A >_\diamond C)\,\&\,(B >_\diamond C)$

Note that, because we do not derive $>_\diamond$ compositionally, $\diamond$-SDA is not simply a special case of SDA. Nonetheless, $\diamond$-SDA is very much in the spirit of SDA itself, and plausibly supported by many of the same intuitive considerations that support SDA.

Semantically, we assume that *might*-counterfactuals existentially quantify over the very same domain that *would*-counterfactuals universally quantify over.

(S1) $[\![\mathsf{A} >_\diamond \mathsf{C}]\!] = \{w \mid R^w \cap [\![\mathsf{A}]\!] \cap [\![\mathsf{C}]\!] \neq \emptyset\}$

Surprisingly, this imposes severe constraints on the range of acceptable meanings for disjunction, ruling out the possibility that a disjunction like A or B has a set of possible worlds as its meaning.

Fact 3. *Assume (S1), the reflexivity of R and the validity of both* SDA *and* $\diamond$-SDA*. Then disjunction is not propositional.*

4 Alternatives

Given our incompatibility results, the prospects for reconciling SDA and CEM might appear bleak. We now turn to strategies for dealing with this tension. Our first attempt is inspired by the alternative semantics for conditionals developed in [1]. In alternative semantics, sentence meanings are not propositions, but instead sets of propositions (or 'alternatives'). A disjunction A or C presents both of A and C as alternatives. That is, $[\![\mathsf{A}\text{ or }\mathsf{B}]\!] = \{[\![\mathsf{A}]\!], [\![\mathsf{B}]\!]\}$. Disjunction contributes a set of propositions as its meaning. SDA can be validated by letting the conditional operate on each alternative in this set.

Our main idea is to derive the meaning of the conditional from an underlying propositional conditional operator $>$–the 'proto-conditional'–which maps a pair of propositions to a new proposition. The proto-conditional regulates the behavior of the conditional $\gg$ when the antecedent is not an alternative. It also helps determine how $\gg$ behaves when its antecedent denotes a non-trivial sets of alternatives. We illustrate this for the case in which $[\![\mathsf{A}]\!]$ denotes a set of propositions.

(S2) $[\![\mathsf{A} \gg\!> \mathsf{C}]\!] = \bigcap \{[\![>]\!](\mathbf{A}, [\![\mathsf{C}]\!]) \mid \mathbf{A} \in [\![\mathsf{A}]\!]\}$

To simplify a bit more, suppose the set of propositions in $[\![\mathsf{A}]\!]$ is $\{\mathbf{B}_1, ..., \mathbf{B}_j\}$ denoted by the sentences $\mathsf{B}_1, ..., \mathsf{B}_j$. Then $\mathsf{A} \gg\!> \mathsf{C}$ is true just in case each of the conditionals $(\mathsf{B}_1 > \mathsf{C}), ..., (\mathsf{B}_j > \mathsf{C})$ is true. In other words, the alternative sensitive conditional is a generalized conjunction of a series of protoconditionals, distributed over the antecedent alternatives.[6] To recycle one of our early examples, the truth-conditions of *Hiro or Ezra* $\gg\!>$ *puzzle* demand the truth of both: *Hiro* $>$ *puzzle* and *Ezra* $>$ *puzzle*.

Before showing how this framework can engage our collapse results, we must make some bookkeeping adjustments. Once we access the higher type of sets of propositions, we need a route connecting them back with propositional meanings. Without such a route, we would not be able to make sense of logical consequence. Furthermore, and relatedly, (S2) does not provide for non-disjunctive antecedents without such a bridge.

We address this problem in a somewhat non-canonical way (for the canonical approach, see [17]). Start by defining the conditional operator polymorphically. That is, let $\gg\!>$ either take a proposition or a set of propositions as input. When it takes a proposition as input, it applies $>$; otherwise, it universally quantifies over alternatives.

(S3) $$[\![\mathsf{A} \gg\!> \mathsf{C}]\!] = \begin{cases} [\![\mathsf{A} > \mathsf{C}]\!] & \text{if } [\![\mathsf{A}]\!] \subseteq W \\ \bigcap \{[\![>]\!](\mathbf{A}, [\![\mathsf{C}]\!]) \mid \mathbf{A} \in [\![\mathsf{A}]\!]\} & \text{otherwise.} \end{cases}$$

Next, we invoke an explicit existential closure operator !. Just like the conditional, we can define our closure operator polymorphically. When $[\![\mathsf{A}]\!]$ is a proposition, ! has no effect on A. But when $[\![\mathsf{A}]\!]$ is a set of propositions, ! takes the union of all of the A alternatives.

(S4) $$[\![!\mathsf{A}]\!] = \begin{cases} [\![\mathsf{A}]\!] & \text{if } [\![\mathsf{A}]\!] \subseteq W \\ \bigcup [\![\mathsf{A}]\!] & \text{otherwise.} \end{cases}$$

Then an argument is valid just in case the closure of the conclusion is true whenever the closure of all the premises are true.

(S5) $$\mathsf{A}_1, \ldots, \mathsf{A}_n \models \mathsf{C} \text{ iff } \bigcap_{i \in [1,n]} [\![!\mathsf{A}_i]\!] \subseteq [\![!\mathsf{C}]\!]$$

This proposal guarantees that disjunction behaves as classically as possible. Since entailment is only sensitive to the closed form of a sentence, we know that *or* satisfies both disjunction introduction and proof by cases.

In this framework, $[\![(\mathsf{A} \text{ or } \mathsf{B}) \gg\!> \mathsf{C}]\!] = [\![\mathsf{A} > \mathsf{C}]\!] \cap [\![\mathsf{B} > \mathsf{C}]\!]$, regardless of what $>$ means. This evidently guarantees that SDA is valid. Whether CEM is valid depends on the choice of proto-conditional $>$. Suppose, following [26], we interpret $>$ in terms of a selection function f that, given a world w and proposition $\mathbf{A}$, returns the unique closest world to w where $\mathbf{A}$ holds.

$$[\![\mathsf{A} > \mathsf{C}]\!] = \{w \mid f(w, [\![\mathsf{A}]\!]) \in [\![\mathsf{C}]\!]\}$$

Then CEM is valid for $\gg\!>$ when the antecedent is not disjunctive.[7] Furthermore, it is a simple corollary of our negative results that there is no non-trivial choice of proto-conditional that

[6] For an implementation of the same idea in inquisitive semantics, with a similar purpose to the one we have here, see [6] and [7].

[7] $\models (\mathsf{A} \gg\!> \mathsf{C}) \vee (\mathsf{A} \gg\!> \neg\mathsf{C})$ iff $[\![!(\mathsf{A} \gg\!> \mathsf{C}) \vee (\mathsf{A} \gg\!> \neg\mathsf{C})]\!] \subseteq W$. But $[\![(\mathsf{A} \gg\!> \mathsf{C}) \vee (\mathsf{A} \gg\!> \neg\mathsf{C})]\!]$ is the set containing $[\![\mathsf{A} > \mathsf{C}]\!]$ and $[\![\mathsf{A} > \neg\mathsf{C}]\!]$, so its closure is the set of worlds where one of these conditionals holds. Since either C or $\neg\mathsf{C}$ is guaranteed to hold at $f(w, [\![\mathsf{A}]\!])$, this last is guaranteed.

validates CEM for disjunctive antecedents. Specifically, CEM fails whenever some alternatives guarantee C and some guarantee $\neg\mathsf{C}$.

We summarize the two signature properties of the semantics above in a single statement.

Fact 4. *For any operator* $>$, $(A \textit{ or } B) > C \models (A > C)\ \&\ (A > C)$.
For any operator $>$, *if* $>$ *validates* CEM, *then* $>$ *validates* CEM *for any* A *not containing or.*

This approach dodges our first two results because those rely on applying CEM to a disjunctive antecedent, and then applying simplification. By blocking CEM for disjunctive antecedents, both proofs are blocked. The current proposal embodies a conservative response to our collapse result: it validates exactly the instances of CEM that do not lead to trouble when combined with SDA.

The problem, however, is that the motivation for CEM does not appear to discriminate against disjunctive antecedents. For instance, (16) and (17) sound equivalent in just the same way that (5) and (6) do

(16) I doubt that if you had slept in or goofed off, you would have passed.

(17) I believe that if you had slept in or goofed off, you would have failed.

Similarly, we observe a duality effect with disjunctive antecedents under *no* and *every*. As before, (18) and (19) appear equivalent.

(18) No student would have succeeded if he had goofed off in class or partied the night before the exam.

(19) Every student would have failed if he had goofed off in class or partied the night before the exam.

By restricting CEM, the analysis renounces these predictions.

Turning to *only if*, we saw that CEM is quite useful in deriving the meaning of *only if* conditionals compositionally from the interaction of *only* and conditionals. Our question now is whether *only if* conditionals with disjunctive antecedents imply their converses.

(20) The flag flies only if the King or Queen is home.

(21) If the flag flies, then the King or Queen is home.

(22) The flag flies if the King or Queen isn't home.

It is clear that (20) does imply (21), just as we saw earlier that (9) implied (10). This is a problem for the analysis above, which denies CEM for conditionals with disjunctive antecedents. For, again, a natural way to predict this entailment is through the idea that *only* negates alternatives, and that (22) is an alternative to the conditional in (20). But if CEM fails for disjunctive antecedents, then the negation of (22) will not imply the contraposition of (21), which is essential in [11]'s account.

Summing up: with alternative semantics, we can enforce SDA while restricting the validity of CEM to non-disjunctive antecedents. However, this restriction is not justified in light of the justification of CEM. For this reason, we now turn to another strategy for avoiding collapse.

5 Homogeneity

We might approach things from the opposite angle: instead of taking an arbitrary conditional and forcing the validity of SDA, we might force the validity of CEM.

5.1 Homogeneity presuppositions

The instrument that yields this result is the theory of homogeneity presuppositions. Homogeneity presuppositions have been invoked to explain certain otherwise problematic variants of excluded middle for plural definites (see for example [11]). In that context, the problem starts with the observation that predications involving plural definites, like (23), plausibly license inferences to universal claims like (24).

(23) The cherries in my yard are ripe.

(24) All the cherries in my yard are ripe.

If some but not all cherries are ripe, one would not be in a position to assert (23). Furthermore, plural definites plausibly exclude the middle. That is, the following sounds like a logical truth:

(25) Either the cherries in my yard are ripe or they (=the cherries in my yard) are not ripe.

If someone were to utter (25), they would sound just about as informative as if they had made a tautological statement (although you might learn from it that they have cherries in their yard). The problem is that, starting with (25) and exploiting entailments like the one from (23) to (24) as well as standard validities for disjunction, we can reason our way to (26):

(26) Either all the cherries in my yard are ripe or all the cherries in my yard are not ripe.

That seems puzzling: did we just prove from logical truths and valid inferences that my yard cannot have some ripe cherries and some non-ripe ones? Of course, something must have gone wrong. The homogeneity view of plural definites explains what that is: first, plural definites carry a presupposition of homogeneity: *the F's are G's* presupposes that the *F*'s are either all *G*'s or all not *G*'s. If this presupposition is satisfied, their content is that all *F*'s are *G*'s. The sense in which (25) sounds tautological is that it cannot be false if its homogeneity presupposition is satisfied. Similarly, the sense in which (23) entails (24) is that if the presupposition of (23) is satisfied and (23) is true, (24) cannot fail to be true. But even if we exploit these to deduce (26), we do not have license us to claim that (26) is valid: our reasoning did not discharge the homogeneity presupposition.

5.2 Forcing CEM via homogeneity

A treatment of CEM using homogeneity presuppositions [11] allows that there may be more than one relevant world where the antecedent of a conditional is true. The key idea is that $\mathsf{A} > \mathsf{C}$ presupposes that C is true at all of the relevant worlds where A is true, or false at all of them. The A-worlds must be "homogeneous" with respect to the consequent.

We generalize the proposal of [11] by reformulating the theory without any appeal to quantification over worlds. Instead, we take an arbitrary conditional operator $>$, and enrich it with homogeneity presuppositions to create a new conditional, $\cdots>$.

(S6) $[\![\mathsf{A} \cdots> \mathsf{C}]\!](w)$ is defined only if $[\![\mathsf{A} > \mathsf{C}]\!](w) = 1$ or $[\![\mathsf{A} > \neg\mathsf{C}]\!](w) = 1$.
If defined, $[\![\mathsf{A} \cdots> \mathsf{C}]\!](w) = [\![\mathsf{A} > \mathsf{C}]\!](w)$.

To talk about SDA and CEM, we also need appropriate assumptions about $\neg$ and $\vee$. These connectives must allow homogeneity presuppositions to project in the right way. To this end, we assume that $[\![\neg\mathsf{A}]\!](w)$ is defined only if $[\![\mathsf{A}]\!](w)$ is defined; if defined, $[\![\neg\mathsf{A}]\!](w) = 1 - [\![\mathsf{A}]\!](w)$.

As for disjunction we assume that $[\![\mathsf{A} \vee \mathsf{B}]\!](w)$ is defined only if $[\![\mathsf{A}]\!](w)$ and $[\![\mathsf{B}]\!](w)$ are defined; if defined, $[\![\mathsf{A} \vee \mathsf{B}]\!](w) = max([\![\mathsf{A}]\!](w), [\![\mathsf{B}]\!](w))$.

Finally, to get predictions about our collapse results, we need a definition of consequence. The leading candidate for languages involving presuppositions is Strawson-validity [28], [11], [12], [13]. According to this notion, an argument is valid just in case the conclusion is true whenever the conclusion is defined and the premises are true.

(S7) $\mathsf{A}_1; \ldots; \mathsf{A}_n \models \mathsf{C}$ iff $[\![\mathsf{C}]\!](w) = 1$ whenever:

- $[\![\mathsf{A}_1]\!](w); \ldots; [\![\mathsf{A}_n]\!](w)$ are defined.
- $[\![\mathsf{A}_1]\!](w) = 1$ and $\ldots$ and $[\![\mathsf{A}_n]\!](w) = 1$.
- $[\![\mathsf{C}]\!](w)$ is defined.

The first important result is that CEM is valid regardless of the choice of proto-conditional. The key result, however, is that any proto-conditional $>$ that validates SDA induces a new conditional $\cdots\!>$ that also validates SDA. Indeed, this is not unique to simplification.

Fact 5. *(i) For any operator $>$, $\models (A \cdots\!> C) \vee (A \cdots\!> \neg C)$; (ii) For any operator $>$, if $>$ validates SDA, then $\cdots\!>$ validates SDA.*

We now have a completely general recipe for validating both SDA and CEM. But have we avoided the bad consequences we claimed should follow? For example, is it the case that for any operator $>$ that validates SDA, $\cdots\!>$ collapses to the material conditional? The answer to both questions is "no". There are many choices of protoconditional for which $\cdots\!>$ is not trivial. A first example is if we let $>$ be a generic strict conditional. To see how this theory avoids triviality, let us look at the semantic correlates of some of the entailments we used in the proof of our first collapse result. The first step of the proof of Fact 1 corresponds to this semantic fact: (27) is a logical truth.

(27) $[(\mathsf{A} \vee \neg\mathsf{A}) \cdots\!> \mathsf{C}] \vee [(\mathsf{A} \vee \neg\mathsf{A}) \cdots\!> \neg\mathsf{C}]$

Although (27) is true whenever defined, it is quite difficult for it to be defined. Given our account of $\vee$, the definedness of $(\mathsf{A} \vee \neg\mathsf{A}) > \mathsf{C}$ is equivalent to the requirement that either $R^w \subseteq [\![\mathsf{C}]\!]$ or $R^w \subseteq [\![\neg\mathsf{C}]\!]$. One of C and $\neg\mathsf{C}$ must be necessary at w (in the relevant sense of necessity) for (27) to be defined.

Now, the reasoning connecting the first two steps of our proof also has a matching semantic fact: (28) entails (29).

(28) $[(\mathsf{A} \vee \neg\mathsf{A}) \cdots\!> \mathsf{C}] \vee [(\mathsf{A} \vee \neg\mathsf{A}) \cdots\!> \neg\mathsf{C}]$

(29) $[(\mathsf{A} \cdots\!> \mathsf{C} \ \& \ \neg\mathsf{A} \cdots\!> \mathsf{C}) \vee (\mathsf{A} \cdots\!> \neg\mathsf{C} \ \& \ \neg\mathsf{A} \cdots\!> \neg\mathsf{C})]$

This holds because if (27) is defined, then the domain R^w uniformly consists of C-worlds or it uniformly consists of $\neg\mathsf{C}$-worlds. Either way, (29) must be true.

Despite the validity of (27) and the entailment from (27) to (29), (29) is not itself valid. The definedness conditions of (29) are laxer than those of (27): for this reason (29) has a much better shot at being false. For instance (29) is false in a model that contains two worlds w and v with w verifying A and C and v verifying $\neg\mathsf{A}$ and $\neg\mathsf{C}$. But such a model does not impugn the validity of (27) under Strawson entailment, because its disjuncts are undefined.

In broad strokes, an instance of transitivity—in particular, one of the form $\models \mathsf{A}$, $\mathsf{A} \models \mathsf{B}$, therefore $\models \mathsf{B}$—fails for Strawson entailment [25]. This is possible because $\models \mathsf{A}$ only requires that A be true if defined; meanwhile, $\mathsf{A} \models \mathsf{B}$ also holds because the presuppositions of A are

essentially involved in guaranteeing the truth of B. But $\models$ B fails because here we are not allowed to assume that the presuppositions of A are satisfied. The same diagnosis applies to our second impossibility result. The first step of the proof claims the validity of $[(\mathsf{A} \text{ or } \mathsf{B}) > C] \vee [(\mathsf{A} \text{ or } \mathsf{B}) > \neg \mathsf{C}]$. The argument establishes that this claim entails IAT. However, the validity of IAT does not follow for a parallel reason to the one we uncovered in discussing the first result.

6 Synthesis

We argued that a generic strict conditional $>$ can validate both SDA and CEM, when enriched with homogeneity presuppositions. Here, however, we must take care. The resulting theory validates SDA, but invalidates $\Diamond$-SDA. That is, the analogue of simplification of disjunctive antecedents for *if ... might ...* fails to be preserved. This is a problem because $\Diamond$-SDA sounds no less plausible than SDA itself.

To fully validate simplification, we propose a synthesis of our two tools. In particular, we suggest that the English conditional recruits *both* alternatives and homogeneity presupposition. To signal this fact, we now introduce the new connective $\cdots\!\!\rightarrowtail\!\!\gg$. Start with any conditional meaning. Then apply the alternative sensitive enrichment from (S5). The resulting semantics validates both SDA and $\Diamond$-SDA, but invalidates CEM for disjunctive antecedents. To validate CEM unrestrictedly, enrich this conditional with homogeneity presuppositions.

More precisely, given an arbitrary proto-conditional $>$, we characterize $\cdots\!\!\rightarrowtail\!\!\gg$ by the clauses:

(S8a) If $[\![\mathsf{A}]\!] \subseteq W$, then $[\![\mathsf{A} \cdots\!\!\rightarrowtail\!\!\gg \mathsf{C}]\!](w)$ is defined only if $[\![\mathsf{A} > \mathsf{C}]\!](w) = 1$ or $[\![\mathsf{A} > \neg \mathsf{C}]\!](w) = 1$.
If defined, $[\![\mathsf{A} \cdots\!\!\rightarrowtail\!\!\gg \mathsf{C}]\!] = [\![\mathsf{A} \rightarrowtail\!\!\gg \mathsf{C}]\!] = [\![\mathsf{A} > \mathsf{C}]\!]$.

(S8b) Otherwise, $[\![\mathsf{A} \cdots\!\!\rightarrowtail\!\!\gg \mathsf{C}]\!](w)$ is defined only if either $[\![>]\!](\mathbf{A})([\![\mathsf{C}]\!])(w) = 1$ for every $\mathbf{A} \in [\![\mathsf{A}]\!]$, or $[\![>]\!](\mathbf{A})([\![\mathsf{C}]\!])(w) = 0$ for every $\mathbf{A} \in [\![\mathsf{A}]\!]$.
If defined, $[\![\mathsf{A} \cdots\!\!\rightarrowtail\!\!\gg \mathsf{C}]\!] = [\![\mathsf{A} \rightarrowtail\!\!\gg \mathsf{C}]\!] = \bigcap\{[\![>]\!](\mathbf{A})([\![\mathsf{C}]\!]) \mid \mathbf{A} \in [\![\mathsf{A}]\!]\}$.

Crucially, there are choices of proto-conditional for which the recipe does not yield a collapsing conditional. In particular, a natural option for the proto-conditional is the Lewisian variably strict conditional. The underlying Lewisian operator allows that there may be multiple worlds where the antecedent is true that are relevant to the evaluation of the consequent. Then the conditional that results from applying the procedure above is doubly homogenous. First, the conditional presupposes that the antecedent alternatives either all guarantee the consequent, or all guarantee the consequent's negation. Second, for each antecedent alternative, the conditional presupposes that either all of the relevant worlds where that alternative holds are worlds where the consequent is true, or they are all worlds where the consequent is false. Perhaps surprisingly, this theory more or less has already been developed and endorsed, for somewhat different reasons, in [24].

References

[1] Luis Alonso-Ovalle. *Disjunction in Alternative Semantics*. PhD thesis, UMass Amherst, 2006.

[2] Stephen Barker. Conditional excluded middle, conditional assertion and *Only if*. *Analysis*, 53:254–261, 1993.

[3] David Butcher. An incompatible pair of subjunctive conditional modal axioms. *Philosophical Studies*, 44(1):71–110, 1983.

[4] Fabrizio Cariani and Simon Goldstein. Conditional heresies. manuscript.

[5] Fabrizio Cariani and Paolo Santorio. *Will* done better: Selection semantics, future credence, and indeterminacy. *Mind*, forthcoming.

[6] Ivano Ciardelli. Lifting conditionals to inquisitive semantics. In *Proceedings of SALT*, volume 26, pages 732–752, 2016.

[7] Ivano Ciardelli, Linmin Zhang, and Lucas Champollion. Two switches in the theory of counterfactuals: A study of truth conditionality and minimal change. June 2017.

[8] Brian Ellis, Frank Jackson, and Robert Pargetter. An objection to possible-world semantics for counterfactual logics. *Journal of Philosophical Logic*, 6:355–357, 1977.

[9] Kit Fine. Critical notice of *Counterfactuals*. *Mind*, 84(335):451–458, 1975.

[10] Kit Fine. Counterfactuals without possible worlds. *Journal of Philosophy*, 109(3):221–246, 2012.

[11] Kai von Fintel. Bare plurals, bare conditionals, and *Only*. *Journal of Semantics*, 14:1–56, 1997.

[12] Kai von Fintel. Npi-licensing, strawson-entailment, and context-dependency. *Journal of Semantics*, 16(1), 1999.

[13] Kai von Fintel. Counterfactuals in a dynamic context. In Michael Kenstowicz, editor, *Ken Hale: a Life in Language*. The MIT Press, 2001.

[14] Kai von Fintel and Sabine Iatridou. If and when 'if'-clauses can restrict quantifiers. manuscript, Massachusetts Institute of Technology, 2002.

[15] Jim Higginbotham. Linguistic theory and Davidson's program in semantics. In Ernie Lepore, editor, *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, pages 29–48. Basil Blackwell, 1986.

[16] Nathan Klinedinst. Quantified conditionals and conditional excluded middle. *Journal of Semantics*, 28(1):149–170, 2011.

[17] Angelika Kratzer and Junko Shimoyama. Indeterminate pronouns: The view from japanese. In Y. Otsu, editor, *The Proceedings of the Third Tokyo Conference on Psycholinguistics*, Tokyo, 2002.

[18] Sarah Jane Leslie. *If*, *unless* and quantification. In R.J. Stainton and C. Viger, editors, *Compositionality, Context and Semantic Values: Essays in Honour of Ernie Lepore*, volume 85 of *Studies in Linguistics and Philosophy*, pages 3–30. Springer Netherlands, 2009.

[19] David Lewis. *Counterfactuals*. Blackwell, 1973.

[20] David Lewis. Possible-world semantics for counterfactual logics: a rejoinder. *Journal of Philosophical Logic*, 6:359–363, 1977.

[21] Donald Nute. Counterfactuals and similarity of words. *The Journal of Philosophy*, 72(21):773–778, 1975.

[22] Donald Nute. Conversational scorekeeping and conditionals. *Journal of Philosophical Logic*, 9(2):153–166, 1980.

[23] Greg Restall. *An Introduction to Substructural Logics*. Routledge, 2000.

[24] Paolo Santorio. Alternatives and truthmakers in conditional semantics. *Journal of Philosophy*, 2017.

[25] Timothy Smiley. Mr. Strawson on the traditional logic. *Mind*, 76(301):347–385, January 1967.

[26] Robert Stalnaker. A theory of conditionals. *American Philosophical Quarterly*, pages 98–112, 1968.

[27] Robert Stalnaker. A defense of conditional excluded middle. In William L. Harper, Robert Stalnaker, and Glenn Pearce, editors, *IFS: Conditionals, Belief, Decision, Chance and Time*, volume 15, pages 87–104. Springer Netherlands, 1981.

[28] P. F. Strawson. *Introduction to Logical Theory*. Methuen, London, 1952.

[29] A. S. Troelstra and Helmut Schwichtenberg. *Basic Proof Theory*. CUP, 2000.

[30] J. Robert G. Williams. Defending conditional excluded middle. *Nous*, 44(4):650–668, 2010.

Referentially used descriptions can be conditionalized*

Eva Csipak

Konstanz University
eva.csipak@uni-konstanz.de

Abstract

In this paper I discuss referentially used DPs such as *Alex's spouse*. I argue that they typically contribute not-at-issue content in the sense of Potts 2005 and Simons et al. 2010, in particular when they occur with adnominal *if*-clauses, such as *Alex's spouse, if they ever got married* or *some soccer player, if that's what he is*. I propose a multi-dimensional analysis, proposing that on the truth-conditional dimension, they simply refer to an individual, while on the non-truth-conditional dimension, they have a modal meaning.

1 Introduction

Since at least Donnellan 1966 [1] it is well-known that utterances containing definite descriptions can make true claims even if the description itself fails to denote, as illustrated in (1).

(1) A (pointing to a man in the room): Alex's spouse is having a good time.
B: Yes, you are right, but they are not married.

B seems to be agreeing with A's claim that the individual is having a good time, while at the same time contesting that the description *Alex's spouse* contains that individual. Donnellan calls this the *referential* use of a definite description, and contrasts it with the *attributive* use. Crucially, speakers are willing to judge A's statement in (1) as true in the context even though it should technically suffer from presupposition failure. Compare this to the attributive use in (2).

(2) A: The owner of this building is rich.
B: #Yes, you are right, but the building is not owned by anyone.

Without knowing exactly what individual *the owner of this building* refers to, A can still have reason to believe that her utterance in (2) is true. Unlike in (1), A does not have any particular individual x in mind that she is referring to. It is not possible for B to at the same time agree that what A said is true, but disagree about the choice of predicate used to refer to *the owner of the building*.

Certain indefinites also have such dual uses, see (3) and (4).

(3) Context: A and B are surveilling a bar. A is working as a bartender while B is watching from a secret room via hidden camera. A sporty-looking person wearing a soccer uniform has just ordered from A.
A: Some soccer player just ordered a beer!
B: Yes, you are right, but that guy is not a soccer player. He's our suspect.

*I gratefuly acknowledge funding by DFG RU 1614 'What If?', project P2. For helpful comments, I thank María Bíezma, Ryan Bochnak, Cleo Condoravdi, Regine Eckardt, Kai von Fintel, Irene Heim, Sven Lauer, Louise McNally, Doris Penka, Maribel Romero, Antje Rumberg, Viola Schmitt, Katrin Schulz, and Sarah Zobel. Errors are my own.

Parallel to (1), A in (3) has an individual in mind that she is referring to with the expression *some soccer player*. Again there is the intuition that A has said something true about that individual, even if *some soccer player* is not true of the individual. There is also a non-referential, quantificational use of those indefinite determiners, as in (4).

(4) A: I have never seen a soccer player.

In (4), A does not need to have a particular individual in mind. In fact, if she had a particular person in mind that she has never seen before, (4) would be a distinctly odd way to express this.[1]

There is a large body of literature surrounding these phenomena which is essentially debating whether these uses are systematically semantically different (referential uses versus attributive and quantificational uses), or whether we can derive the differences in interpretation from some pragmatic mechanism.

For definite descriptions, Donnellan himself is the first of many to argue that there is a semantic ambiguity between referential uses and attributive uses. Kripke 1977 [6] and much subsequent literature argues for a pragmatic account instead. For indefinites, a semantic ambiguity approach is proposed e.g. by Fodor & Sag 1982 [2], while Kratzer 1998 [5] argues that the referent is identified as the value of a choice function which is supplied by the context (i.e., a mostly pragmatic mechanism).

In this paper, I argue for a semantic ambiguity approach. I propose that referentially used definite DPs identify their referent, and the semantic content they themselves provide is only added as not-at-issue material in the sense of Simons et al. 2010 [10]. Referentially used indefinites do provide at-issue material, as do attributively used definite descriptions and quantificational DPs. Both types of referentially used descriptions are firmly in the not-at-issue dimension, however, once they are modified by an adnominal *if*-clause. Adnominal *if*-clauses are *if*-clauses as in (5) which seem to modify the nominal, rather than the matrix proposition as a whole.

(5) Alex's spouse, if they ever got married, is having a good time.

Intuitively we take the clause *if they ever got married* to modify *Alex's spouse*, not *Alex's spouse is having a good time*. This will be discussed in more detail in section 2, where I also argue that referentially used definite DPs provide not-at-issue content. I provide an account of adnominal *if*-clauses that supports this view.[2] There is a recent proposal by Frana that I compare to my proposal in section 3. Section 4 discusses several open questions.

2 Referentially used DPs, at-issueness, and adnominal *if*-clauses

Donnellan 1966 [1] argues that definite descriptions are systematically ambiguous, and that we need to distinguish between 'speaker's reference' and 'semantic reference'. This was picked up

[1] Note that numerals also fall into this category of indefinite determiners that can be used referentially.

(i) Two men have proposed to Alex in the last 24 hours (but I won't tell you who).

[2] I will stick with the term *if*-clause rather than *antecedent* as a way to help us remind ourselves that these are not standard conditionals; the matrix clause does not receive a modal interpretation in this account.

and defended by Stalnaker 1970 [11] who proposed the following semantics, in the formalization given by Heim 2011 [4].

(6) $[\![the_{ref-stal}\alpha]\!]^{c,i} = \iota x$ [the speaker in c presupposes $\lambda i'.[\![\alpha]\!]^{c,i'}(x)$]

That is, a definite description *the* α denotes a unique individual x, and the speaker presupposes that x counts as an α. It seems reasonable to translate Heim's *the speaker in c presupposes* into an epistemic modal.

(7) $[\![the_{ref}\alpha]\!] = \iota x\ [\forall w' \in Best(\bigcap f(w)):\ \alpha(x)$ in $w']$

Again a definite description *the* α denotes a unique individual x, but now the speaker presupposes that in all her (best) epistemically accessible worlds, x counts as an α.

On such a view, the main lexical content contributed by the definite description is essentially a presupposition, and the only non-presuppositional content of the definite description is 'ιx', i.e., the unique individual that is identified.

Let us now consider adnominal *if*-clauses like (5) in more detail. We had the intuition that the *if*-clause modifies the content of the definite description. That means that on this view, it modifies the presupposition. A set of literature that provides tools to enable presuppositions to participate in compositional semantics is the literature on not-at-issue content, and in particular on multi-dimensional semantics. We therefore set out to show that the content of the definite description is not-at-issue, in hopes of employing the tools of multi-dimensional semantics to account for the meaning of adnominal *if*-clauses.

Following Potts 2005, 2007 [8, 9] and Simons 2010 [10] I assume that at-issue material can be easily negated and denied, whereas not-at-issue material cannot. Referentially used definite descriptions provide not-at-issue content. Consider the following contrast.

(8) A (pointing to a man on the dance floor): I hear Alex's spouse has texted you.
B: No, that's not true. #Alex and that guy are not married.
B': No, that's not true. He has called me.

B's denial can only mean that the individual A is pointing at has not texted B (which is at-issue); what it cannot mean is that that individual is not Alex's spouse. This shows that the contribution of *Alex's spouse* is not part of the at-issue meaning. Now consider the following example, where *the building manager* is used attributively.

(9) A: I hear the building manager has texted you.
B: No, that's not true. It was a neighbour.
B': No, that's not true. She has called me.

Here A is using *the building manager* attributively; A does not need to have a precise idea of who this individual is, and B's denial can target the definite description, showing that its contribution is at-issue.

Interestingly we observe a contrast between referentially used definite descriptions which are not-at-issue and referentially used indefinite descriptions which are at-issue, as illustrated in (10).

(10) A (pointing to a man at the bar): Some soccer player has just arrived.
B: No, that's not true. That's my priest.

Clearly B can deny the content of the referentially used indefinite DP *some soccer player* here. But consider (11), where A adds an adnominal *if*-clause.

(11) A (pointing to a man at the bar): Some soccer player, if that's what he is, has just arrived.
B: No, that's not true. #That's my priest.
B': No, that's not true. He's been there the whole time.

Suddenly B can no longer deny the material provided by the DP. This is surprising. To look for an explanation, we turn to adnominal *if*-clauses in more detail.

2.1 The properties of adnominal *if*-clauses

We first establish that adnominal *if*-clauses do not have the same interpretation as standard conditionals. Adnominal *if*-clauses, unlike hypothetical or biscuit conditionals, can only occur parenthetically or postposed; they cannot occur preposed. This is illustrated in (12) with the definite referential DP.

(12) a. Alex's spouse, if they ever got married, just started dancing.
b. Alex's spouse just started dancing, if they ever got married.
c. ≠ If they ever got married, Alex's spouse just started dancing.

The only interpretation that (12-c) can receive is an odd one where there is some kind of rule in place such that if the couple gets married, Alex's spouse is forced to dance. But this is not the same reading that is available in the other cases where only *Alex's spouse* is modified: in (12-c) the entire sentence is interpreted as a hypothetical conditional.

We have already seen that the adnominal *if*-clause in (11) only contributes not-at-issue material. This is also the case when it occurs with definite DPs, as in (13).

(13) A: Alex's spouse just started dancing, if they ever got married.
B: No, that's not true. #Alex's spouse didn't start dancing if they ever got married.

The antecedents of hypothetical and biscuit conditionals, on the other hand, do contribute at-issue content, as illustrated in (14) and (15), respectively. In both cases, B can target the conditional relation between antecedent and consequent and deny it.

(14) A: We will go to the park if the weather is good.
B: No, that's not true. We will not go to the park if the weather is good.

(15) A: There is pizza in the fridge if you are hungry.
B: No, that's not true. There is no pizza in the fridge if I'm hungry.

Thus examples (13) – (15) show that adnominal *if*-clauses differ systematically from different types of conditionals in that they provide not-at-issue content. This suggests that they are less similar to conditionals, and perhaps more similar to other types of clauses that contribute not-at-issue material, such as non-restrictive relative clauses like (16).

(16) A: Alex's spouse, who is a keen dancer, has just arrived.
B: No, that's not true. #He does not like to dance.

The relative clause *who is a keen dancer* contributes not-at-issue information. We can model this in a multi-dimensional semantics, and to make things easier for readers unfamiliar with more recent works, I use the system provided by Potts 2005 [8]. Note however that in cases where the material needs to interact with both truth-conditional and non-truth-conditional

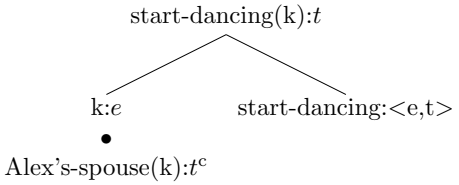
material, we need to assume a hybrid dimension as has been proposed by McCready 2010 [7] and subsequent authors.

2.2 The proposal

[8] proposes to treat both non-restrictive relative clauses and 'supplements' in the following way (the simplified parsetree in (18) models both (17-a) and (17-b)). The bullet operator • separates the two dimensions, and non-truth-conditional types are indicated by a superscript ^{c}.

(17) a. Kim, Alex's spouse, just started dancing.
b. Kim, who is Alex's spouse, just started dancing.

(18)

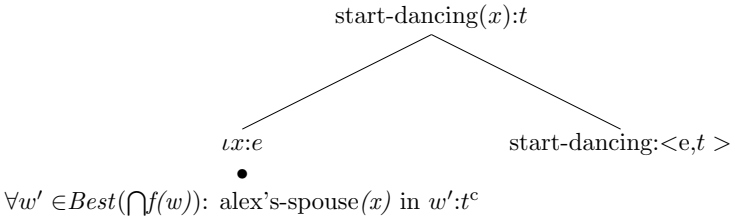


On the truth-conditional dimension, *Kim* refers to an entity which is available for functional application with the predicate *just started dancing* in the familiar way. On the non-truth-conditional dimension, the predicate *Alex's spouse* is applied to *Kim*. The overall sentence meaning is computed as the proposition that Kim started dancing, and the additional not-at-issue information that Kim is Alex's spouse. This corresponds to the intuition we wanted to model.

Applying this mechanism to referentially used definite descriptions yields the following.

(19) A (pointing): Alex's spouse just started dancing.

(20)

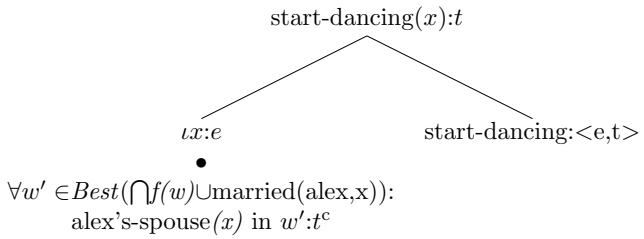


The unique individual that the definite description refers to is the truth-conditional component of *Alex's spouse*. This predicts that even if that individual is not Alex's spouse, the at-issue proposition will not suffer from presupposition failure. It can simply be true or false, depending on whether or not the individual started dancing. On the not-at-issue level, the speaker signals that in all her best epistemically accessible worlds, x (the referent selected by *Alex's spouse* in the actual world) is Alex's spouse in that world.

We can now simply add an adnominal *if*-clause.

(21) A (pointing): Alex's spouse, if they ever got married, just started dancing.

(22)

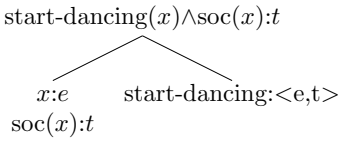


The *if*-clause restricts the epistemic modal provided by the referentially used definite description, and the not-at-issue content is now that in all the speaker's best epistemically accessible worlds *where Alex is married to x*, x is Alex's spouse. This leaves open the possibility that there are among the best epistemically accessible worlds those where the two are not married. In those worlds, no prediction is made as to whether the individual is Alex's spouse. But importantly, independent of whether the actual world is one where they are married, the truth-conditional contribution of *Alex's spouse* is to refer to a particular individual, just like it did in (20).

In the case of referentially used indefinites, we could follow e.g. Fodor & Sag 1982 and assume that a referentially used indefinite *some* α also identifies an individual x, and conveys $\alpha(x)$. Note that we saw that indefinites convey at-issue information, but that when they combine with an adnominal *if*-clause, their contribution is no longer at issue.

(23) A (pointing): Some soccer player just started dancing.

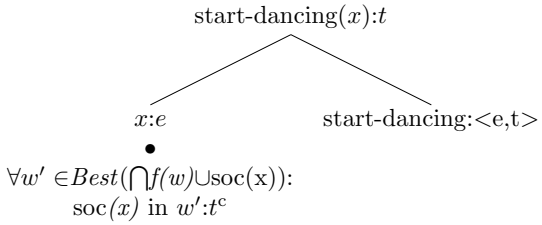
(24)



Here, *some soccer player* identifies an individual x and predicates two things of x: x is a soccer player, and x just started dancing. Both are at-issue (they be the target of negation and denial etc.). Once we add in the adnominal *if*-clause, the only at-issue content that *some soccer player* provides is that it refers to a salient individual.

(25) Some soccer player, if that's what he is, just started dancing.

(26)



That is, the fact that x is a soccer player is no longer at-issue material. Recall from (10) that we can no longer deny x being a soccer playier by saying 'that's not true!'. Instead, in the not-at-issue we now have the information that in all the speaker's best epistemically accessible worlds where x is a soccer player, x is a soccer player.

3 An alternative account

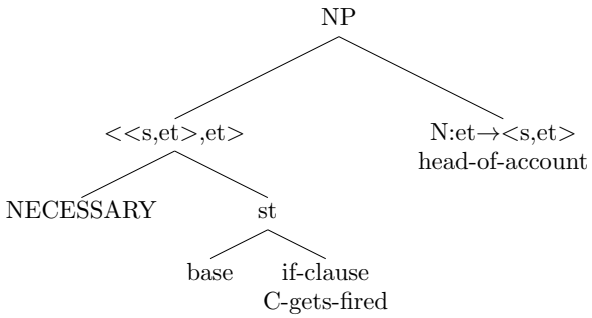
There is a recent proposal for adnominal conditionals by Frana 2017 [3]. The examples she considers are more modal in nature than the ones we have considered so far, like (27).

(27) The head of accounting, if Campbell gets fired, is currently working for a competitor in London.

The definite description *the head of accounting* identifies an individual who at speech time is intuitively not in the denotation of the NP: the individual works for a competitor, so cannot be the head of accounting at the speaker's company. It is understood that that individual will perhaps become the head of accounting once Campbell gets fired, which may or may not happen. Thus the speaker of (27) is in a different epistemic state from the speakers we have considered so far: so far, the speakers have typically used definite descriptions in contexts where they were uncertain whether the individual they were referring to was in the extension of the NP. The speaker of (28) has no such uncertainty: she believes that the individual is certainly *not* in the extension of *the head of accounting* at speech time.

The analysis that Frana proposes is the following. There is a covert modal necessity adjective, and it is this adjective that the *if*-clause restricts, as in (28).

(28)



In her account, the *if*-clause attaches below the NP level, which means that she does not have to treat definite and indefinite DPs separately. However, because of this low attachment, she also predicts that adnominal *if*-clauses should be able to appear in quantifiers, even universal ones. This is not the case, see (29).

(29) a. #Every student, if we can call him that, cheated on the exam.
b. #Every ex-husband, if they ever got married, brought Alex a rose.

She also predicts the material provided by the *if*-clause to be at-issue. Since her examples are different from the ones we have seen, let us consider her example.

(30) A: The head of accounting, if Campbell gets fired, is currently working for a competitor in London.
B: No, that's not true. #She will be the CFO./#Campbell won't get fired, just demoted.
B': No, that's not true. She is working in Paris.

It is clear from B's failed denial that the material in the *if*-clause is not-at-issue. This needs to be built into the account. One way of doing this would be to simply propose that the covert modal adjective operates on the not-at-issue dimension rather than the at-issue dimension.

However, even then at least one issue remains. Languages like German allow APs to host a lot of very diverse material, for example adverbials and discourse particles (for more data, see Viesel 2015 [12]). Consider the following examples.

(31) der morgen *(notwendige) Arztbesuch
the tomorrow necessary doctor.visit
'the doctor's visit that is necessary tomorrow'

(32) die ja *(notwendige) Neubesetzung
the PART necessary new.placement
'the – as we know – necessary replacement'

These examples show that adverbials like *tomorrow* and discourse particles like *ja* are not able to occur in DPs unless an adjective is present. These need not be modal adjectives, but since Frana posits a covert adjective expressing modal necessity, these examples show that *tomorrow* and *ja* are compatible with the overt version of such adjectives.

We expect that any material that can be hosted in an AP that is projected by an overt adjective should also be able to be hosted in an AP projected by a covert adjective. But here, this is not what we find: neither adverbials nor discourse particles can occur with DPs that host adnominal *if*-clauses, see (33) and (34).

(33) #der morgen *(notwendige) Arztbesuch, wenn man es einen Besuch nennen kann
the tomorrow necessary doctor.visit if one a visit call can
intended: 'the doctor's visit, if one can call it that, that is necessary tomorrow'

(34) #die ja *(notwendige) Neubesetzung, wenn Campbell gefeuert wird
the PART new.placement if Campbell fired will
intended: 'the –as we know – necessary replacement if Campbell gets fired'

Notice that the intended meaning is coherent, as is illustrated by the fact that when the *overt* adjective is present, the adverbial and particle are acceptable in (33) and *ja*, respectively. It is only when there is no overt adjective and Frana's covert adjective should take over that the constructions become unacceptable. This means that if one wants to maintain a covert AP account, more needs to be said.

4 Open questions

So far we have not considered whether the content of the adnominal *if*-clause has an influence on its acceptability. In fact this does seem to play a role: only certain types of *if*-clauses can appear adnominally. For example, it is not possible to express a law-like dependency between DP and *if*-clause.

(35) #That aggressive man, if he drinks, works at a café.

The phrase *that aggressive man, if he drinks* in (35) cannot mean 'this unique x' and non-truth-conditionally express 'in all the best epistemically accessible worlds where the guy drinks, he is an aggressive man'. But this is what we predict. There has to be a further restriction: only those *if*-clauses can restrict definite descriptions that express a necessary condition for the DP to hold of x. This is illustrated with the following two contexts.

(36) Context: During the Olympics, a runner A has qualified for the final round. The finalists all start in a group. The fastest qualifying time was run by runner B, who

was 3 seconds faster than A. In principle, if everybody else runs as fast as during the qualifying time and A runs 4 seconds faster, A wins.
Newscaster, talking about A: #The winner, if she beats her qualifying time by 4 seconds, is getting ready.

In this context, it is not *necessary* for A to beat her time by four seconds in order to win. If everybody else runs much slower this time, A could run the same speed and still win. We observe: the sentence is odd. Contrast that with the following example.

(37) Context: During the Olympics, a long jumper A has qualified for the final round. The finalists all take turns, and so far everyone has done exceptionally badly. In fact, A is the last person to jump and can win by jumping only 3,00 meters, which is considered very easy.
Newscaster, talking about A: The winner of the 2016 Olympics, if she makes this final jump farther than 3 meters, is getting ready.

In this context, jumping 3 meters is a necessary condition for A to be the winner: A has to jump at least 3,00 meters to win. If she jumps less far or does not jump at all, she cannot win. In this example, it also happens to be a sufficient condition that A jump 3,00 meters to be the winner, but this is not required. Recall example (27), repeated here as (38).

(38) The head of accounting, if Campbell gets fired, is currently working for a competitor in London.

We understand that while it is a necessary condition for Campbell to be fired in order for the salient individual to become head of accounting, it is by no means a sufficient condition. This restriction that the *if*-clause needs to express a necessary condition is not currently predicted by our account, and one might wonder if there is a pragmatic principle that can predict this, or if we need to derive it in some other way.

Another interesting challenge is the treatment of proper names. It seems clear that we can modify proper names with adnominal *if*-clauses.

(39) A: Alex, if that was his name, came to the party.
B: No, that's wrong. His name was Bob.

Notice that *Alex* remains at-issue content and can be targeted by B's denial. While this in itself does not posit a technical problem for multi-dimensional semantics (certain elements can be used on more than one dimension at once), it is surprising that definite DPs seem be fully on the not-at-issue dimension, whereas proper names appear in both dimensions. This requires further thought.

A final puzzle is that even non-referential uses of DPs seem to be able to be modified by adnominal *if*-clauses. Consider the following examples.

(40) Context: A and B are investigating the death of Smith. It seems that another person was involved, but it is unclear if this person maliciously attempted to kill Smith, or if it was an accident. At present A and B do not know the identity of this person.
A: The perpetrator, if we can call him that, has left some DNA.

A's use of *the perpetrator* seems to be attributive, in that the speaker does not have a particular individual in mind.

We can even get quantificational readings of indefinites to occur with adnominal *if*-clauses,

as in (41).

(41) If a student – if you can still call him a student – cheats, every professor will get fired. (adapted from Kratzer 1998)

Here the speaker is expressing uncertainty about whether someone who behaves in an unstudentlike way (for example, someone who cheats) can be called a student, without having a particular individual in mind. Examples like these suggest that we have to rethink our claim that what is left on the truth-conditional level is an individual of type e. On the other hand, if we adopt a quantificational analysis of *a student* in (41), more needs to be said about why other quantifiers such as *every* seem to always be unacceptable with adnominal *if*-clauses.

The proposal presented here provides some new empirical evidence to shed light onto the debate on referentially used DPs. It argues that at least when combining with adnominal *if*-clauses, referentially used DPs contribute not-at-issue content.

References

[1] Keith Donnellan. Reference and definite descriptions. *The Philosophical Review*, 75(3):281 – 304, 1966.

[2] Janet Fodor and Ivan Sag. Referential and quantificational indefinites. *Linguistics and Philosophy*, 5:355 – 398, 1982.

[3] Ilaria Frana. Modality in the nominal domain. In Ana Arregui, Maria Luisa Rivero, and Andres Salanova, editors, *Modality across syntactic categories*, pages 49 – 69. Oxford University Press, 2017.

[4] Irene Heim. Definiteness and indefiniteness. In Klaus von Heusinger, Claudia Maienborn, and Paul Portner, editors, *HSK 33.2*, pages 996 – 1025. de Gruyter, 2011.

[5] Angelika Kratzer. Scope or pseudo-scope? are there wide-scope indefinites? In S Rothstein, editor, *Events in Grammar*, pages 163 – 196, Dordrecht, 1998. Kluwer.

[6] Saul Kripke. Speaker's reference and semantic reference. *Midwest Studies in Philosophy*, 2, 1977.

[7] Eric McCready. Varieties of conventional implicature. *Semantics and Pragmatics*, 3(8):1 – 57, 2010.

[8] Christopher Potts. *The logic of conventional implicatures*. Oxford University Press, 2005.

[9] Christopher Potts. The expressive dimension. *Theoretical Linguistics*, 33(2):165 – 198, 2007.

[10] Mandy Simons, Judith Tonhauser, David Beaver, and Craige Roberts. What projects and why. In Nan Li and D Lutz, editors, *Proceedings of SALT 20*, pages 309 – 327, 2010.

[11] Robert Stalnaker. Pragmatics. *Synthese*, 22:272 – 289, 1970.

[12] Yvonne Viesel. Discourse structure and syntactic embedding. In Thomas Brochhagen, Floris Roelofsen, and Nadine Theiler, editors, *Proceedings of the 20th Amsterdam Colloquium*, pages 418 – 427, 2015.

Counterfactual Double Lives*

Michael Deigan

Yale University
michael.deigan@yale.edu

Abstract

Expressions typically thought to be rigid designators can refer to distinct individuals in the consequents of counterfactuals. This occurs in counteridenticals, such as "If I were you, I would arrest me", as well as more ordinary counterfactuals with clearly possible antecedents, like "If I were a police officer, I would arrest me". I argue that in response we should drop rigidity and deal with *de re* modal predication using something more flexible, such as counterpart theory.

We often talk about what would have happened had things been otherwise.

(1) If I had taken I-95, I would have been caught in traffic.

Even though this describes the merely possible scenario(s) of me taking the highway, it nevertheless seems that I myself, inhabitant of the actual world, appear in it. The sentence, if I've uttered it, is about *me*. How best to understand this kind of *de re* modal discourse has long been a matter of controversy, particularly the question of 'transworld identity'.[1] Should we understand the 'I' of the consequent as referring to a person numerically identical to the actual speaker of the utterance? Following the work of Kripke (1980) and Kaplan (1989b), this is the orthodox view. Or does it pick out possible individuals not identical to me, but ones related to me by a certain counterpart relation? This is the position taken by Lewis (1973), among others.

In this paper I raise a new problem for treating *de re* modal talk in terms of transworld identity. We lead counterfactual double lives—the same person can make multiple *distinct* appearances in the same counterfactual scenario. A transworld identity theorist cannot make sense of this, since it violates the logic of identity. If this is right, the fact that counterpart theorists don't have the same problem with double lives is a big advantage for their theory.

I begin with a discussion of counteridenticals, a kind of counterfactual in which counterfactual doubling often plays a prominent role, and spell out the argument that making sense of this doubling requires that we drop the transworld identity view. Then I respond to two objections to the argument. In doing so I show that counterfactual doubling occurs outside counteridenticals, in more mundane counterfactuals with clearly possible antecedents.

1 Double Lives in Counteridenticals

Counteridenticals are counterfactuals whose antecedents involve apparent statements of identity between individuals who are actually distinct.

(2) a. If I were you, I'd bring an umbrella.
b. If I were Shaq, I'd be 7 ft tall.

*Thanks to Josh Knobe, Tim Williamson, and Ken Winkler for discussion of this and a related paper, and to Bill Lycan for discussion and encouragement at a much earlier stage.

[1] For an overview, see Mackie and Jago (2013).

c. If you were me, you would have done the same thing.

These are commonly used in giving advice, as in (2-a). But as we can see in (2-b) and (2-c), that's not all they can do.[2]

Counteridenticals have received some attention,[3] but given their various interesting properties, not nearly as much as they deserve. In this paper I will be focusing on just one of these interesting properties, leaving a fuller discussion. The property that I will be focusing on is that in counteridenticals, individuals can have counterfactual double lives, seeming to make multiple distinct appearances in the same counterfactual scenario. Contrast the following.

(3) a. If I were you, I would be pleased with myself.
b. If I were you, I would be pleased with me.

In (3-a) the speaker is talking about how, in a certain scenario, one individual would relate to herself. In this case, she would be pleased with herself; no other individuals need be involved. But in (3-b), this is not so. The scenario being discussed is one in which one individual, referred to by the *I* of the consequent, is pleased with another, distinct individual, referred to by the *me* of the consequent. The speaker of (3-b) seems to lead a kind of counterfactual double life.

And it's not just that *I* and *me* can pick out different individuals. In counteridenticals, different occurrences of *my* can be about possessive relations to different individuals, and even different occurrences of *I* can come apart in reference.

(4) a. If I were you, I would proofread my paper.
b. If I were you, I would make sure that I proofread my paper.

In (4-a), the recommendation might be that the addressee proofread the addressee's paper, or it might be that she proofread the speaker's paper. In (4-b), the ambiguity of the second *I* of the consequent and the *my* give rise to at least three readings: recommendation that the addressee (i) make sure the addressee proofreads the addressee's paper, (ii) make sure the addressee proofreads the speaker's paper, and (iii) make sure the speaker proofreads the speaker's paper. So each of the two counterfactual lives of the speaker can be picked out by an *I* in the consequent of a counteridentical, sometimes even when the *I* is in the same surface position of the sentence.[4]

The double life phenomenon in non-advisory uses of counteridenticals.

(5) a. If I were Shaq, I could look down on me.
b. If I were Shaq, I would have dunked on me.

We also see it with other referring expressions, including other pronouns and names, which are typically thought to be rigid. Counteridenticals with the speaker as subject are the most familiar, but others are not unnatural.

[2] This productivity makes them unlikely candidates for being treated as idioms.

[3] Going back at least to Goodman's 1946 lecture on counterfactuals reprinted in Goodman (1983). See also Lewis (1973, p. 43), Lakoff (1970), Lakoff (1996), Reboul (1996), Arregui (2007), and Thomas (2008). There is also recent, more detailed work—Carina Kauf (2017) and Alex Kocurek (forthcoming)—of which I was unaware until after having written this paper. Kocurek independently makes an argument against the standard Kripkean view, and develops a version of the Lewisian account to which I am sympathetic.

[4] Interestingly, it's difficult to use *I* to refer to the person usually picked out by *me* if the *I* is not embedded further in some way. This can happen, though, when the thematic role of the subject is non-agentive. A prisoner might say to his captor: "If I were you, I'd have been released already".

(6) a. I'm not sure why you're angry. If you were me, you wouldn't have waited for you either.
b. I'm not sure why he's angry. If he were her, he would have told him the same thing.
c. I'm not sure why John is angry. If John were Mary, John would have borrowed his car without asking, too.

So the argument below will apply not only to first-person pronouns, but also to second-person pronouns (see (6-a)),[5] third person pronouns (see (6-b)), and names (see (6-c)).

2 The Problem for Kripke-Kaplanian Orthodoxy

The problem for transworld identity is simple: the two counterfactual 'lives' of the speaker are distinct, so by the transitivity and symmetry of identity, they can't both be identical to the same thing, since that would imply they are identical with each other. But the standard Kripke-Kaplan theory of indexicals implies that what *I* designates and what *me* designates in the above examples *are* both identical to the speaker. Let's see why.

Kripke (1980) argues that with respect to metaphysical modality, names in natural languages act like constants, designating the same individual in any circumstance of evaluation.[6] This contrasts with descriptions like 'the number of planets' which may vary in reference across different circumstances of evaluation. To use Kripke's terminology, names are thought to be *rigid designators*, which designate the same individuals in all possible worlds. More carefully and specifically: *weak* de jure *rigid designators*, as opposed to strongly or *de facto* rigid. That is, *weakly* rigid because a name is only claimed to denote the same individual in all worlds *where the relevant individual exists*, with no further claim about what it denotes in worlds where that individual doesn't exist. And *de jure* because names are claimed to be rigid in some sense by their very semantics, or 'by stipulation', rather than by some reliance on descriptions which in fact hold necessarily, like "the smallest prime".

Kaplan (1989b) showed how we can extend this kind of view to indexicals like *I* and *me* by making a "sharp distinction between contexts of use and circumstances of evaluation" (Kaplan 1989b, p. 495). Once a context of use provides some individual (in the case of *I* and *me*, the speaker in that context), that same individual will be referred to in all circumstances of evaluation.[7] That names and indexicals are weak *de jure* rigid designators with respect to metaphysical modality is by now the orthodox view, though by no means a consensus. But even those who depart from it usually aim to maintain rigidity in less direct ways, e.g. by use of rigidifying operators.[8]

[5] We also see this in an amusing counteridentical from Wodehouse (2007, p. 246):

> 'Why? Be reasonable, Bertie. If you were your aunt, and you knew the sort of chap you were, would you let a fellow you knew to be your best pal tutor your son?'
>
> This made the old head swim a bit, but I got his meaning after awhile, and I had to admit that there was much rugged good sense in what he said.

[6] With respect to epistemic modality, it is generally thought that names are *not* rigid. See Fitting and Mendelsohn (1998) and Holliday and Perry (2014).

[7] In fact, Kaplan was arguing for what he took to be the logically stronger claim that indexicals are *directly referential*, in the sense that the semantic rules for them "*do not provide a complex which together with a circumstance of evaluation yields an object. They just provide an object*" (Kaplan 1989b, p. 495). Though not all rigid designators, even *de jure* ones, will be directly referential, he thinks all directly referential expressions will be rigid (Kaplan 1989a, p. 571). If this is right and if I'm right that indexicals are not rigid, it will follow that they are not directly referential, either.

[8] See, among others, Geurts (1997), Elbourne (2005), Maier (2009), Hunter (2013), and Fara (2015).

We need to introduce one more part of the standard view contributed by Kaplan: that metaphysical modals are not *monsters*—they are not operators which shift contextual parameters.[9] This means that in a given context, an indexical will have the same intension regardless of whether it is inside or outside the scope of metaphysical modal operators.

Now let's look at how the standard Kripke-Kaplan story goes for a normal metaphysical modal attribution using a first-personal pronouns, such as in a counterfactual.[10] Suppose someone—call him Donald—utters the following.

(7) If I were a police officer, I would arrest someone.

Because Donald is the speaker, an unembedded occurrance of *I* will refer to him. And because *I* is rigid, any occurrence of *I* will refer to Donald in every circumstance of evaluation. And because there are no monsters here, the *I* in the consequent will also have to refer to Donald, again rigidly. So (7) will be true iff in the closest worlds where Donald is a police officer, Donald arrests someone in those worlds. This seems right, as far as it goes. 'Score: 1' for the orthodoxy.

But what happens when we try to run the Kripke-Kaplan account on a counteridentical which involves double lives? We get a contradiction. Start with the following contrast.

(8) a. If I were you, I would arrest myself.
b. If I were you, I would arrest me.

We note that in (8-b), as is clear when contrasted between (8-a), the consequent describes situations in which one person arrests another—no self-arresting need be involved. It's not the case that the *I* and *me* of the consequent must corefer; indeed, it seems that they must not corefer. But assuming unembedded occurrences of *I* and *me* both refer to the speaker in the context, and that there are no monsters, and that *I* and *me* are rigid designators, then the occurrances of *I* and *me* in the consequent of (8-b) must both refer to the speaker of the context in any circumstance of evaluation. Given the necessity of identity, the *I* and *me* of the consequent must corefer. But this is a contradiction—we've said that it's not the case that the corefer.

In the course of deriving our contradiction, we relied on a few substantive planks of the classical Kripke-Kaplan platform. They are: (i) Rigidity of *I* and *me*, (ii) No Monsters, and (iii) Necessity of Identity. To account for double lives in counteridenticals, I believe we should drop (i). But it's worth noting that getting rid of any one of these will undermine the claim that *de re* modal sameness is one of transworld identity in anything like the sense that Kripke and others believe. Giving up rigidity means, straightforwardly, that when considering counterfacutal scenarios, the referent of *I* is not in general going to be identical to the referent in the actual world. If we say that what goes on in counterfactuals is some sort of monstrous context-

[9] Famously, Kaplan makes the much broader claim that there can be no monsters in natural languages. Following the work of Schlencker (2003) and others, this is now widely thought to be incorrect—it seems there are monsterous operators in at least some natural languages. Santorio (2010) argues that epistemic modals, even in English, are monsters. But nobody, as far as I know, claims metaphysical modals are monsters, which is the relevant position here.

[10] Nothing in the argument will turn on the details of any specific analysis of counterfactuals. For concreteness, I'll use the familiar Stalnaker-Lewis picture that a counterfactual is true when the closest worlds where the antecedent is true are worlds where the consequent is true.

I will be assuming, though, that counterfactuals and metaphysical modality are tightly linked, and that what goes for the treatment of *de re* attributions in one also goes for the others. I will be appealing to data from counterfactuals and drawing conclusions about metaphysical modality in general. I do not take the required connections to be particularly controversial, but they can be motivated in different ways. Most directly, if we follow the restrictor analysis of Kratzer (1986), counterfactuals in fact involve an metaphysical necessity modal. See Williamson (2007) for motivation from a different perspective.

parameter-shifting, we might be able to technically hold onto rigidity, but we'll have to give up the view that a non-counterfactual use of the same indexical uttered in the same context will refer to a same individual. And to make the monster solution work, we'd need an account of how the contextual parameters get shifted; I suspect we'd have to end up relying on a kind of accessibility relation that will be much like a counterpart relation. And finally, I know of no way of dropping (iii) while maintaining anything like the thesis of transworld identity for *de re* modal predication. The standard way of making sense of 'contingent identity' after Kripke's work has been to appeal to counterpart theory and give up rigidity of the relevant terms.[11]

So I suggest we drop rigidity and treat *de re* modal predication using something more flexible. Counterpart theory is the obvious option. Happily, it has no trouble in allowing for counterfactual double lives. We just need to allow for multiple counterpart relations to be used in the same sentence, indexed to different occurrences of the referring terms. This is the path Lewis (1973, p. 43) takes for other reasons, and it turns up in his suggestion for dealing with counteridenticals:

> For a familiar illustration of the need for counterpart relations stressing different respects of comparison, take *'If I were you . . . '*. The antecedent worlds are worlds where you and I are vicariously identical; that is, we share a common counterpart. But we want him to be in *your* predicament with *my* ideas, not the other way around. He should be your counterpart under a counterpart relation that stresses similarity of predicament; mine under a different counterpart relation that stresses similarity of ideas.

This is also compatible with double lives: the *I* and the *me* of the consequent just need to be indexed to different counterpart relations, one relating the speaker to the shared counterpart, another relating her to another individual, e.g., one with a similar causal history to the speaker's.

This is by no means a complete theory, and it would take a lot of work to get it sufficiently into shape.[12] Its flexibility makes it prone to problems of overgeneration. But the task of accounting for counterfactual double lives within a counterpart theoretic framework seems feasible, whereas the Kripke-Kaplanian account seems to rule it out in principle. This seems to me a strong argument in favor of dropping rigid designation and moving to a counterpart theoretic treatment of *de re* modality.[13]

3 Objections and Replies

Orthodoxy tends not to cede easily. Let's consider a couple objections its defenders might make.

3.1 The Impossibility Objection

The impossibility objection goes as follows. The argument from double lives relies on treating counteridenticals as normal counterfactuals with antecedents that are possibly true. But counteridenticals are not normal counterfactuals. They are counterpossibles: counterfactuals

[11] See Gibbard (1975), Lewis (1971), and Schwarz (2013).

[12] In other work, I propose a way of generalizing this theory, then refine it to avoid certain problems. Kocurek (forthcoming) also develops Lewis's proposal.

[13] We should be careful to not overstate the argument's scope. It doesn't undermine the very idea of transworld identity, it just undermines its use in dealing with *de re* modal ascriptions. Indeed, the counterpart-theoretic account of double lives in counterfactuals which I favor is compatible with using a possible worlds semantics with a constant domain. And it's even compatible with there being some expressions, or some uses of some expressions, which rigidly refer to these individuals.

with impossible antecedents. By the necessity of identity (and rigidity of *I* and *you*), I couldn't have been you; to assume otherwise is just to beg the question against the transworld identity theorist. And weird things happen with counterpossibles...

At this point, the objection diverges into two versions. On one major view, counterpossibles are all vacuously true. There are no possible worlds in which the antecedent holds, so it follows that in all of the worlds in which it does hold (all 0 of them), the consequent holds as well. If we accept this, then there's no problem in explaining the possibility of true utterances of any of the sentences in (8) and (6).

What can't be so easily explained is why some counteridenticals seem to be false, like certain utterances of the above sentences, or ones like.

(9) If I were Shaq I'd be 200 ft tall.

Here we must say that utterances of these are all true, and merely seem false. We'll have to rely on some non-semantic explanation of why they seem false.[14] As difficult as this task may be, there's no strong reason yet to think there's a threat to rigidity here.

The second version of the objection could be made by someone who takes counterpossibles to be sometimes substantively true and other times substantively false. There are various ways to go about doing this, each revising the standard view of possible world semantics for counterfactuals in some way. One common strategy is to introduce impossible worlds. If we do this, we might admit that it's true that the referents in the consequent of a double-life counteridentical are distinct, but nevertheless claim we need not drop transworld identity. We say this is just another impossibility that arises in the counterpossible scenario—that something is distinct from itself. We'll need to develop an account of identity and reference in counterpossible scenarios, but it's not clear that there will be any problem with rigidity in doing so, and even if some problem arises, we need not think it should reflect back on rigidity in ordinary counterfactuals dealing only with possible worlds.

The impossibility objection is troubling to the argument from counteridentical double lives. There are various rejoinders to it we might pursue. We might defend the view that the antecedents really are possible, and say that the Kripke-Kaplan view rules this out is just a further problem with the view. Or we might criticize the most plausible versions of the pragmatic or revisionary accounts that could be used to save rigidity, or defend an alternative as doing a better of accounting for various data.[15] Instead of taking of any of these direct routes, though, we can follow an easier path: we can simply make the argument using cases of counterfactual double lives with clearly possible antecedents.

3.2 Double Lives Outside of Counteridenticals

So far we've only seen the phenomenon of counterfactual double lives in the rather strange (but nevertheless common and productive) construction that is the counteridentical. Though this is where the phenomenon is most prevalent, it also appears elsewhere. Consider the following.

(10) If I were a police officer, I would arrest me.

This seems acceptable (and true in some contexts), but is no counteridentical—the antecedent is just an ordinary one, as in

(11) If I were a police officer, I would have a badge.

[14] See Williamson (2015).
[15] This last is the path taken by Kocurek (forthcoming).

But if this is right, then counterfactual double lives appear in counterfactuals with clearly possible antecedents. We can then restate the argument of §2 using these other counterfactuals and thereby totally avoid the impossibility objection.

Besides appearing in counterfactuals, they also can arise in (metaphysical) modal subordination,[16] and the subordinating supposition, again, can be clearly possible.

(12) I could have been a police officer. I would have arrested me for what I just did.

They can also be less blatant than in these cases of distinct referents for *I* and *me*. For example, a wily criminal, discussing the investigation into a crime she herself committed, might say

(13) If I were a detective, I would have solved this crime ages ago.

The criminal makes a double appearance here both explicitly as the detective solving the crime and implicitly as the criminal who the detective discovers did it. To make the double life more apparent, note that she might follow with

(14) I would have realized that only I was capable of getting through the bank's security.

And of course there is no suggestion that it would have been a detective who committed the crime and discovers herself to have done it, as in some psychological thriller.

Counterfactual double lives, then, are more widespread than they first appear, and are not limited to cases with antecedents of questionable possibility. And this means the impossibility objection won't work.

The objector might respond that while the antecedent of (10) *appears* possible, as the antecedent of (11) is, it actually isn't, but is instead a counteridentical in disguise. On this view, we would have to read *a police officer* in a *de re* way, and it would have the same content as our old (8-b) if uttered by the speaker to some particular police officer. If this is right, then the antecedent really is still like the ones before, so the argument would still be subject to the impossibility objection.

I agree that there is such a reading of (10), but I deny that this is the only one. We can perfectly well say the following, forcing a *de dicto* reading of *a police officer*.

(15) If I were a police officer—not any particular actual police officer (for all I know there aren't any police officers left), just if I were some police officer—I would arrest me.

It is not plausible that this has a *de re*, disguised counteridentical reading. Thus the impossibility objection is indeed dodged.

There are some last ditch efforts we can try on behalf of the impossibility objector, though. We could say that even though the antecedent is possible, something impossible is happening in the situation the consequent describes anyways. We could then say these are trivially false (and rely heavily on pragmatics), or take the revisionary approach of appealing to impossible worlds, as before. This suggestion, though, is unappealing for various reasons. Among them the fact that it requires giving up the following principle: if $p > q$ and $\Diamond p$, then $\Diamond q$. Given that, at least on the *de dicto* reading of *a police officer*, the antecedent of (10) is clearly possible in some contexts where an utterance of the sentence would seem true, we should conclude that what the consequent describes is also possible.

Another response would be to say there's some implicit material in the antecedents of counterfactuals with double lives which, when spelled out, reveal them to be impossible. But

[16] On modal subordination, see Roberts (1989).

without some story of what that material is and why we should think it is there, this suggestion is *ad hoc*, and I see little reason to accept it if there are alternatives.

3.3 Descriptive Indexicals?

Another objection, though, does apply to cases of counterfactual lives outside of counteridenticals. The objection is that we've relied on too simple a view of how *I* and *me* and other referring expressions work, and that once we have a truer, more complicated picture, we can hang on to the Kripke-Kaplan orthodoxy, or at least most of it.

According to this objection, indexicals can be used in the normal, *de re* way, but they also have a descriptive use, on which they behave like a contextually supplied definite description which in fact holds of the normal referent, but will refer to whichever individual satisfies the description in other possible worlds. That is, it admits that there are non-rigid uses of indexicals. This, of course, is a real restriction of the standard view, which is usually taken to be fully general.[17] But we might think it's a restriction we should have already made in response to 'descriptive indexicals', which are already fairly well known.[18] These are uses of indexicals like the following.

(16) a. I am traditionally allowed to order whatever I like for my last meal. (Nunberg 1993, p. 20)

b. [*pointing at Pope Francis*] He is usually an Italian. (Elbourne 2013, p. 202)

Here the *I* means something like *the condemned prisoner* and *he* means something like *the pope*. And descriptive indexicals seem to work in counterfactuals.

(17) If we had abolished the electoral college, he [*pointing at Trump*] would be a woman.

The proposal, then, is that at least one of indexicals in a double-life counterfactual is a descriptive indexical, so doesn't have any bearing on the restricted standard theory, which maintains rigidity only for 'plain', non-descriptive uses of indexicals. Thus the objector can maintain that there's no *new* problem for the standard view from counterfactual double lives. And much of the Kripke-Kaplan picture can remain intact, depending on how we react to descriptive indexicals.

This objection fails, though. Counterfactual double lives cannot be successfully treated in terms of descriptive indexicals. Take one of our double life cases, (10), repeated as (18).

(18) If I were a police officer, I would arrest me.

The claim we're considering is that at least one of the pronouns in the consequent is a descriptive indexical. But I don't think we can take either of them to be descriptive indexicals.

It can't be the *I*, because there isn't any contextually salient description that holds of the speaker in the actual context and is used to pick out the referent in other circumstances of evaluation. The obvious candidate is *the police officer*, but this can't work, since it's presumed that the speaker is not in fact a police officer. But what *I* would have to mean if it were being used descriptively is some description which does hold of the actual speaker.

And it can't be *me*, because when we use tricks which force a non-descriptive reading, we still understand the *me* in the same way. Consider what happens when we add nominal appositives with names to the sentences that had descriptive readings of indexicals available.

[17] Though we might follow the strategy of Kripke (1977) and try to treat these as non-literal uses for which semantic reference and speaker reference diverge and thereby maintain the rigid semantics for all occurrances of indexicals, making this version of the objection non-revisionary to the standard view.

[18] For a discussion of recent attempts account for descriptive indexicals, see Sæbø (2015).

(19) a. He, Pope Francis, is usually an Italian.
b. If we had abolished the electoral college, he, Donald Trump, would have been a woman.

The descriptive readings vanish. Where we once had true and non-maxim-violating readings of the sentences which relied on descriptive uses of the indexicals, now there are no such readings to be found. But what happens when we try the same with (18)?

(20) If I were a police officer, I would arrest me, Mike Deigan.

Nothing much happens. It's a bit more explicit, and there will likely be some pragmatic effects of mentioning the name, but there's no blocking of a double-life reading. And if there were, we'd expect the reflexive pronoun to be used instead, so we'd expect this new sentence to sound ungrammatical, but it doesn't. So even when we force a non-descriptive reading of *me*, we still get the same effect, so it can't be a descriptive reading of *me* that is doing the work.

So it's neither a descriptive reading of the *I* or the *me* in (18) that's responsible for the second counterfactual life. I conclude that the effect of counterfactual double lives is not attributable to descriptive uses of indexicals.

We've now considered a couple objections to the argument from counterfactual double lives and found that neither holds much promise for rescuing the Kripke-Kaplan account. If there's no alternative that does better, we may wish to reexamine these or consider other objections to the argument. Luckily, though, it seems that the main alternative approach to *de re* modal predication—the counterpart theoretic approach—has the resources to handle counterfactual double lives.

References

Arregui, Ana (2007). "Being Me, Being You: Pronoun Puzzles in Modal Contexts". In: *Proceedings of Sinn Und Bedeutung* 11.

Davidson, Donald and Gilbert Harman, eds. (1972). *Semantics of Natural Language*. Reidel.

Elbourne, Paul (2005). *Situations and Individuals*. MIT Press.

— (2013). *Definite Descriptions*. Oxford University Press.

Fara, Delia Graff (2015). "Names are Predicates". In: *Philosophical Review* 124.1, pp. 59–117.

Fitting, Melvin and Richard L. Mendelsohn (1998). *First-Order Modal Logic*. Kluwer Academic Publishers.

Geurts, Bart (1997). "Good news about the description theory of names". In: *Journal of Semantics* 14, pp. 319–348.

Gibbard, Allan (1975). "Contingent Identity". In: *Journal of Philosophical Logic* 4, pp. 187–221.

Goodman, Nelson (1983). *Fact, Fiction, and Forecast*. Fourth Edition. First published in 1955. Harvard University Press.

Holliday, Wesley H. and John Perry (2014). "Roles, Rigidity, and Quantification in Epistemic Logic". In: *Trends in Logic, Outstanding Contributions: Johan van Benthem on Logic and Information Dynamics*. Ed. by Alexandru Baltag and Sonja Smets. Springer, pp. 591–629.

Hunter, Julie (2013). "Presuppositional Indexicals". In: *Journal of Semantics* 30, pp. 381–421.

Kaplan, David (1989a). "Afterthoughts". In: *Themes from Kaplan*. Ed. by Joseph Almog, John Perry, and Howard Wettstein. Oxford University Press, pp. 565–614.

— (1989b). "Demonstrativess: An Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives and Other Indexicals". In: *Themes from Kaplan*. Ed. by Joseph Almog, John Perry, and Howard Wettstein. Oxford University Press, pp. 481–564.

Kauf, Carina (2017). "Counterfactuals and (Counter-)Identity: The Identity Crisis of "If I Were You"". MA thesis. Georg-August-Universtät Göttingen.
Kocurek, Alex (forthcoming). "Counteridenticals". In: *Philosophical Review*.
Kratzer, Angelika (1986). "Conditionals". In: *Chicago Linguistics Society* 22, pp. 1–15. Reprinted with revisions in Kratzer (2012, pp. 86–108).
— (2012). *Modals and Conditionals*. Oxford University Press.
Kripke, Saul (1977). "Speaker's Reference and Semantic Reference". In: *Midwest Studies in Philosophy* 2, pp. 255–276. Reprinted in Kripke (2011).
— (1980). *Naming and Necessity*. Harvard University Press. Originally published in Davidson and Harman (1972).
— (2011). *Philosophical Troubles: Collected Papers, Volume 1*. Oxford University Press.
Lakoff, George (1970). "Linguistics and Natural Logic". In: *Synthese* 22, pp. 151–271. Reprinted in Davidson and Harman (1972).
— (1996). "Sorry, I'm Not Myself Today: The Metaphor System for Conceptualizin the Self". In: *Spaces, Worlds, and Grammars*. Ed. by Gilles Fauconnier and Eve Sweetser. University of Chicago Press.
Lewis, David (1971). "Counterparts of Persons and Their Bodies". In: *The Journal of Philosophy* 68.7, pp. 203–211. Reprinted with postscript in Lewis (1983).
— (1973). *Counterfactuals*. Blackwell.
— (1983). *Philosophical Papers*. Vol. I. Oxford: Oxford University Press.
Mackie, Penelope and Mark Jago (2013). "Transworld Identity". In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Summer 2013. URL: https://plato.stanford.edu/archives/fall2013/entries/identity-transworld/.
Maier, Emar (2009). "Proper names trigger rigid presuppositions". In: *Journal of Semantics* 23, pp. 253–315.
Nunberg, Geoffrey (1993). "Indexicality and Deixis". In: *Linguistics and Philosophy* 16, pp. 1–43.
Reboul, Anne (1996). "If I were you, I wouldn't trust myself". In: *Acts of the 2nd International Colloquim on Deixis "Time, space and identity"*, pp. 151–175.
Roberts, Craige (1989). "Modal Subordination and Pronominal Anaphora in Discourse". In: *Linguistics and Philosophy* 12, pp. 683–721.
Sæbø, Kjell Johan (2015). "Lessons from Descriptive Indexicals". In: *Mind* 124.496, pp. 1111–1161.
Santorio, Paolo (2010). "Modals are monsters: on indexical shift in English". In: *Proceedings of SALT* 20, pp. 289–308.
Schlencker, Philippe (2003). "A plea for monsters". In: *Linguistics and Philosophy* 26.1, pp. 29–120.
Schwarz, Wolfgang (2013). "Contigent Identity". In: *Philosophy Compass* 8.5, pp. 486–495.
Thomas, Guillaume (2008). "Proxy counterfactuals". In: *Snippets* 18, pp. 17–18.
Williamson, Timothy (2007). *The Philosophy of Philosophy*. Blackwell Publishing.
— (2015). "Counterpossibles". In: *Proceedings of the 20th Amsterdam Colloquium*. Ed. by Thomas Brochhagen, Floris Roelofsen, and Nadine Theiler, pp. 30–39.
Wodehouse, P. G. (2007). *The Best of Wodehouse: An Anthology*. Everyman's Library. Alfred A. Knopf.

Learning what *must* and *can* must and can mean*

Annemarie van Dooren[1], Anouk Dieuleveut[1],
Ailís Cournane[2] and Valentine Hacquard[1]
[1] University of Maryland, USA
[2] New York University, USA
avdooren@umd.edu, adieulev@umd.edu, cournane@nyu.edu, hacquard@umd.edu

Abstract

This corpus study investigates how children figure out that functional modals like *must* can express various flavors of modality. We examine how modality is expressed in speech to and by children, and find that the way speakers use modals may obscure their polysemy. Yet, children eventually figure it out. Our results suggest that some do before age 3. We show that while root and epistemic flavors are not equally well-represented in the input, there are robust correlations between flavor and aspect, which learners could exploit to discover modal polysemy.

1 Introduction

Almost half of the world's languages have modal forms that express different "flavors" of modality [1]. For instance, English *must* can express both epistemic and deontic necessities, as well as various other "root" (i.e. non epistemic) flavors (e.g. teleological, bouletic): (1) can mean that John is required to eat meat (deontic necessity) or that he is probably a meat eater (epistemic necessity). We use the term "polysemy" atheoretically to refer to this behavior.

(1) John must eat meat.

In this paper, we ask when and how children figure out that modals like *must* are polysemous. To this end, we examine how modality is expressed in speech to and by children: how often it is expressed using *lexical* modals (verbs, adjectives or adverbs like *maybe*), which are typically monosemous, vs. *functional* modals (modal auxiliaries like *must*), which can be polysemous.

Picking up on the flavor polysemy of modals may be challenging for several reasons. First, learners may need to overcome word learning biases [2]: some modals (e.g. *must*) can express different meanings, violating the *principle of contrast*, and different modals can express the same meaning (e.g. *maybe* and *might*), violating the

* We would like to thank our research assistants Joon Lee and Jan Michalowski, the ModSquad @ UMD and audiences at Harvard, Rutgers, and Dubrovnik. Our project is supported by NSF grant #BCS-1551628.

principle of mutual exclusivity. Second, modal flavor may not be obvious from the situational context alone: modals express abstract concepts with no reliable physical correlates to give away the intended flavor. Moreover, the context is often compatible with different flavors (if John *is allowed to* eat meat, he *plausibly* does), and even adults can't always tell the intended flavor [3].

Results from the existing acquisition literature suggest that children may not initially realize that functional modals can be epistemic: they do not produce functional epistemics until age 3, a year after they start producing root flavors ([5], [6], [7], [8], a.o.). This so-called 'epistemic gap' has been argued to reflect a conceptual lag ([9], [4]), or a grammatical lag ([8], [10]). However, children do produce "lexical" epistemics during the epistemic gap (e.g. *maybe*; [11]). This suggests that the epistemic lag is not primarily conceptual, since it is tied only to functional modals [8]. If children are not producing *functional* modals with epistemic flavors at first, have they perhaps not yet realized that these modals can express epistemic modality ([4]:387)? Does this gap arise from properties of the input?

To date, no study has extensively investigated the input, as the focus has been on children's productions. Here, we ask how adults use polysemous modals, and whether root and epistemic flavors are equally well-attested. We show that the way adults use functional modals may obscure their polysemy: epistemic modality is rarely expressed using functional modals. Furthermore, speakers tend to use polysemous modals in a monosemous way. Yet, children eventually figure out modal polysemy. We show that some may do so even before age 3. Given that modal flavor may not be evident from the situational context alone, and that the way speakers use polysemous modals obscures their polysemy, we ask whether cues to the polysemy of modals might come from their syntactic distribution.

We explore in particular correlations between modal flavor and modal syntax that have emerged from the literature on modality, notably in how modals interact with tense and aspect. We focus here on the fact that root and epistemic modals differ in *temporal orientation*: root modals tend to be future-oriented, epistemic modals tend not to be ([12], [13], [14], [15], a.o.). We show that while root and epistemic flavors are not equally well represented in the input, there are clear distributional differences that index these differences in temporal orientation. We sketch how children could exploit these distributional cues to figure out modal flavor, and in turn, modal polysemy.

2 Study

We examined the modal productions of 12 children and their mothers from the Manchester corpus [16], on the CHILDES database [17]. These child-mother dyads were recorded for one hour in play sessions, twice every three-week period, over the course of one year, from age 1;09 to age 3;00. All utterances containing modal words were extracted (81,854 of 564,625 total utterances). We chose this corpus for its density and uniformity of sampling sessions during the so-called epistemic gap period. This allows us to get a more accurate picture of rare early child uses of epistemics

than previous studies, and the uniformity across 12 dyads allows us to generalize observed patterns above and beyond individual differences.

Modals were coded for syntactic category (*functional*: auxiliaries, quasi-auxiliaries; *lexical*: adverbs, adjectives, verbs), as shown in (2), and for flavor (root, epistemic, metaphysical). Note that we do not differentiate amongst root flavors (e.g. ability, teleological, deontic), and leave the question of how children figure out root polysemy for future work. We also coded modal complements for aspect: grammatical (progressive, perfect), lexical (eventive, stative).

(2) Modal lemmas by syntactic category:
Functional Aux = can, could, may, must, should, might, shall, will, would
Quasi-Aux = have to, got to, ought to, supposed to, going to
Lexical V = *epis*: know, think, seem...; *root*: want, order... Adv = *epis:* maybe, probably... Adj = *epis*: sure, certain... *root*: able, capable...

2.1 Input: mothers' production

To get a sense of the kind of modals children are exposed to, we ask how frequently parents express epistemic vs. root modality, and how frequently modality is expressed using functional vs. lexical modals. The results for modal talk by category are summarized in Table 1. We find that for lexical modals, both epistemic and root modality are equally well attested in the input (4.6% of all mother utterances contain a lexical epistemic vs. 3.7% for lexical roots). Functional modals are well-represented in the input: 13% of all mother utterances. Thus, children hear a fair amount of epistemic modal talk, and a fair amount with functional modals. Whatever is responsible for the purported epistemic gap, it is not a lack of exposure to epistemic talk. Example input utterances are given in (3).

Table 1: Modal input per category (12 adults, % of total utterances)

Lexical modality			Functional modality	
epistemic	root	epis/root	epis/root	future
15,750 (4.6%)	12,433 (3.7%)	2,434 (0.7%)	20,528 (6%)	22,661 (6.7%)
30,617 (9%)			43,189 (12.7%)	

(3) Examples of modal utterances from the input
a. Lexical epistemic: Maybe$_{epi}$ there are no trousers. Mother (Ruth 2;00)
b. Functional epistemic: It might$_{epi}$ be cold in Scotland. Mother (Aran, 2;10)
c. Lexical poly: What do we need to draw first then? Mother (Aran, 2;03)

To investigate how well modal polysemy is represented in the input, we focused on functional modals that can express root or epistemic flavors (*can, could, may, must, should, have_to, got_to, supposed_to, ought_to* and *might*), and determined the intended flavor in context for six mothers[1]. Table 2 shows the distribution of root vs. epistemic flavors for each modal. We find that functional modals are overall used much more frequently to express root (92%) than epistemic (8%) modality. This effect is driven by the fact that the most frequent modals (*can, have to*) nearly always express root modality. Our results further show that modals that are in principle

[1] 10% of modals were double-coded, 99% overlap.

polysemous are mostly used monosemously: *can, could, have to, got to, should, supposed to* and *ought to* express root modality more than 90% of the time. *Must* and *may* are more often used with epistemic flavors. *Might* expresses epistemic possibility 65% of the time and metaphysical possibility 35% of the time.

Table 2: Polysemous modals by flavor (6 adults & children)

Modal	ADU Total	ADU Root	ADU Epi	ADU % root	CHI Total	*CHI* Root *(+repetition)*	*CHI* Epis *(+repetition)*	CHI % root
can[2]	5262	5230	32[3]	99%	2004	180 (+27)	1	100%
have to	1024	1020	4	100%	120	113 (+ 7)	0	100%
could	791	718	73	91%	54	39 (+ 11)	4	91%
might	592	205 (meta)	387	35%	66	19 (+ 25) (meta)	15 (+ 8)	46%
got to	522	519	3	99%	176	145 (+ 31)	0	100%
should	338	318	20	94%	18	9 (+ 9)	0	100%
must	199	40	159	20%	40	22 (+ 9)	6 (+ 3)	79%
supposed	111	102	9	92%	8	8	0	100%
ought to	12	12	0	100%	12	12	0	100%
may	12	4	8	33%	6	1 (+ 1)	4	20%
Total	8863	8167	696	92%	2592	547	51	91%

Thus, epistemic and root flavors are not equally well represented in the input, in ways that might make it challenging to see that functional modals can be polysemous[4]: such modals are in practice largely monosemous, and overall, used more frequently for root than epistemic modality. Do young children still pick up on modal polysemy? How well does their own production mirror that of their parents?

2.2 Children's modal production

We examined children's modals to see to what extent they reflect input. We see that children produce a fair amount of functional modals (3% of total utterances), though proportionally less so than their parents. They also produce some lexical epistemics (0.8% of total utterances), though proportionally less so than their parents, and less than they produce lexical root modals. These results are summarized in Table 3. Thus, while young children may be less disposed to express epistemic modality, they do produce some lexical epistemics well before age three.

Table 3: Modal output per category (12 children, % of total utterances)

Lexical modality			Functional modality	
epistemic	root	epis/root	epis/root	future
1,911(0.8%)	7,475 (3.3%)	1,003 (0.4%)	5,389 (2.4%)	2,305 (1%)
10,389 (4.6%)			7,694 (3.4%)	

[2] For child *can*, we used a random sample of 208 occurrences out of 2004 total occurrences.

[3] *Can* only has root interpretations in the adult grammar, except under negation. Half of the adults' epistemic *cans* were under negation. The other half were in questions, such as "Where can it be?". Such uses may be circumstantial (*possibility given the circumstances*) or epistemic (*possibility given the evidence*). In cases where it is difficult to tease apart epistemic from root modality, we erred on the side of epistemics.

[4] We hope to test this claim using computational modeling in the future.

We examined the functional modals produced by 6 children[5] (age 2;0-2;11), and find that children *do* produce epistemic modals, albeit much less frequently than roots (Table 2). The epistemic modals children produce are those that are most often used epistemically by adults (*might, must, may*). Yet, we see that flavor distribution for children's functional modals does not mirror that of their parents, particularly in the case of *must*, which is mostly used with epistemic flavors for adults (80%), but with root for children (79%), suggesting that they have a root bias, at least in production.

We further examined the first occurrences of various modals. For all children, *can* appears before other modals, in line with previous findings. Yet, children's first uses of *might*[6]*, could,* and *must* with an epistemic flavor occur before age 3, as does *maybe*. Furthermore, three out of the six children use *must* with *both* epistemic and root flavors before age 3 (5).

(5) a. Epistemic: it must be some of dolly's hair. (Aran, 2;09)
b. Root: I must get crane. (Aran, 2;02)

Our results show that children produce lexical and functional epistemics before age 3, suggesting that the epistemic gap from the literature may be due to the lower sampling density of previous studies. Further, at least some children use some polysemous modals with both root and epistemic flavors, suggesting they have worked out their polysemy.

2.3 Summary

Our corpus results show that the way speakers express modality might make it challenging to see that functional modals can be polysemous. Children, however, do pick up on modal polysemy, maybe even earlier than has been assumed in the literature. How do children figure it out? We hypothesize that children make use of distributional cues to hone in on the kinds of flavors their modals express. In this paper, we focus specifically on potential aspectual cues, building on insights from the semantic literature.

3 Aspectual cues to modal polysemy

How do children figure out that the same modal words can express different modal flavors? Paying attention to just the situational context may not settle the matter as possibilities do not have reliable physical correlates and the context is often compatible with different types of possibilities. Finally, the way speakers use modals does not provide ample opportunities to observe that the same modals express different flavors. Yet, young children work it out. We explore the possibility that to do so, children exploit temporal-aspectual cues that differentiate modal flavors.

[5] *Might*: 33% double-coded, 95% overlap. Other modals: 25% double-coded, 95% overlap.

[6] The first uses of *might* reported in previous literature appear with so-called 'physical predicates' [11] like *fall*, and are likely metaphysical. In our corpus, epistemic *might*, like metaphysical *might*, before age 3.

3.1 Modal flavors and Temporal Orientation

While context plays a big role in determining modal flavor [18], the availability of various modal flavors seems to be constrained by their interactions with tense and aspect ([19], [20], [21], [22], a.o.). In particular, many argue that root and epistemic modals differ in the kinds of ***temporal orientation (TO)*** they can have: root modals tend to be future-oriented (the time of the prejacent event has to follow the time at which the modal is evaluated), but epistemics tend not to be: they can have Past or Present TO ([12], [13], [23], [15], [14], a.o.).

TO arises from combinations of lexical and grammatical aspect in the modal's prejacent, as illustrated in (6). Progressive aspect results in present TO: in (6b) the possibility is about a concurrent run; Perfect aspect results in past TO: in (6c) the possibility is about a past run. In the absence of an overt grammatical aspectual, TO depends on the lexical aspect of the prejacent: eventives trigger future TO (6a) (or present TO with a habitual reading), statives lead to present TO (6d) (future TO is possible for instance with an adverbial like *later*). Root flavors are available only when a future TO is possible, i.e. in the absence of an aspectual operator in the prejacent (6a), and more easily with eventives than stative prejacents.

(6) a.	John may run.	Future TO, Present TO (habitual)	epis, root
b.	John may be running.	Present TO	epis, *root
c.	John may have run.	Past TO	epis, *root
d.	John may be home.	Present TO, %Future TO	epis, %root

These constraints, if well exemplified in the input, could provide useful cues to the learner: data like (6b) and (6c), with a progressive or perfect in the prejacent, which only allow non future TO, could hint that the modal expresses epistemic modality. Present-oriented stative prejacents might hint at epistemic flavors as well.

3.2 Aspectual cues in the input

Turning first to *grammatical aspect*, we expect epistemics, but not roots, to take complements with perfect and progressive aspects. This is what we find (Table 4): functional modals have embedded aspect 11% of the time when epistemic (7), but less than 1% of the time when root. All root modals with an embedded perfect had a counterfactual interpretation (8). Note that we do find a few cases of roots with embedded progressive, with *supposed to*, *should* and *got to* (9), which suggest that root modality is occasionally non future oriented.

Table 4: Grammatical aspect (6 adults)

	Epistemic (n=696) *(n% of total epistemics)*	Root (n=8167) *(n% of total roots)*
Progressive	14 (2.0%)	28 (0.3%)
Perfect	65 (9.3%)	40 (0.5%)
Total	79 (11.4%)	68 (0.8%)

(7) a. Because if it's got wet it might not be working properly (Mother, Aran 2;08)
b. Somebody *must have locked* the door to the post office (Mother, Aran 2;08)

(8) You *should have eaten* it at dinnertime. (Mother, Anne 2;03)
(9) You're not *supposed to be eating* the stethoscope. (Mother, Ruth 2;02)

As for *lexical aspect*, we expect epistemics to combine more readily with statives, and roots with eventives, if stative prejacents typically trigger present TO, and eventives future TO. We classified prejacents consisting of predicates that lacked an overt aspect using classic tests from [24]. Note that some predicates sometimes seem stative and sometimes eventive (perception verbs, *think*, and *have*). Because we did not want to commit to a particular view on how children interpret them, and the cues that these verbs provide may in fact be complex, we treated them all uniformly. The numbers reported in Table 5 show what the proportions are like when we treat them as eventives; the numbers in parentheses show the proportions if we were to treat them as statives.

Table 5: Lexical aspect (3 adults)

	Epistemic (n=316) *(n% of total epistemics)*	Root (n=4394) *(n% of total roots)*
+ stative	83% (91%)	6% (26%)
+ eventive	17% (9%)	94% (74%)

We see that stativity of the prejacent correlates with flavor: if we treat perception predicates as eventives, we see that roots take mostly eventives (94%) and epistemics take mostly statives (83%). If we treat them as statives, the link between stativity and epistemic modality is even more accentuated (91%). Further details with a breakdown per modal is provided at http://ling.umd.edu/~hacquard/project_modality.html

In sum, for functional modals, our corpus data show clear distributional differences between flavors, in terms of the aspectual properties of the modals' prejacents. These differences could provide useful cues to the learner that the modals can express different flavors. In the next section, we sketch how a syntactic bootstrapping account might work.

4 Bootstrapping modal polysemy from aspectual cues

According to the *syntactic bootstrapping hypothesis* ([25], [26], a.o.), children hone in on a word's meaning by exploiting principled links between its meaning and its syntactic distribution. This learning strategy may be critical for abstract meanings that lack clear physical correlates ([27], [28], [29], [30]). Modal meanings may be difficult to observe, aspect morphemes may be easier. If children expect modal flavors to correlate with temporal orientation, and different aspect combinations to trigger different TOs, they could exploit aspectual cues to work out modal flavor. In particular, non future TO may cue in the learner that a modal is epistemic. This bootstrapping account makes two crucial assumptions: 1) the link between TO and

flavor is principled; 2) children are able to pick up on and exploit aspectual cues. We turn briefly to each of these assumptions.

4.1 Motivating constraints on modal flavor and TO

The literature on modality argues that the link between modal flavor and TO is principled. Several authors propose that the link between root flavors and future TO is due to a general constraint which prevents vacuous uses of modals, e.g., Condoravdi's *Diversity Condition (DC)* [12], which requires that there are worlds in the modal base where the prejacent p is true and worlds where it is not.

(10) ***DC:*** For worlds w, times t, common ground cg, modal base MB, and property P, there is a w ∈ cg and w', w'' ∈ MB (w, t) such that (t, w', P) and ¬(t, w'', P).

Because epistemic and root modals differ in the kinds of facts relevant for their modal bases, the DC applies differently, in ways that interact with TO. Epistemic modals take an epistemic modal base, which picks out worlds compatible with a body of knowledge; root modals take a circumstantial modal base, which picks out worlds compatible with some circumstances[7].

Condoravdi first introduced the DC to explain why metaphysical modals (which [15][8] assume are a subset of *circumstantial*, i.e. root, modals) disallow non future TO. The past and the present are "settled", hence the same facts hold throughout metaphysically accessible worlds (or more generally, worlds compatible with the circumstances). This means that, when a root modal has present or past TO, the worlds of the modal base are uniform, and thus cannot differ with respect to whether p holds or not. The future, on the other hand, is not settled, hence the worlds of the modal base can differ w.r.t. p with future TO. The DC however allows epistemics with non future TO: an epistemic modal base picks out worlds compatible with a body of knowledge; what we know about the past or the present may leave some uncertainty about the truth of the prejacent p, hence the modal base can have both p and not p worlds.

The DC thus explains why root modals can only have future TO, but epistemics need not. Whether epistemics allow future TO is a matter of debate: some argue that epistemics disallow future TO because of the incompatibility of epistemicity with the uncertainty of the future [14]. If learners expect modal meanings to be governed by something like the DC, perhaps because of a more general expectation about non vacuity, they may be able to use TO to hone in on modal meanings.

4.2 Children's understanding of aspectual cues

Assuming that the links between aspectual properties and modal flavors are principled, can children make use of aspectual cues? Are they sensitive to aspectual distinctions, and do they exploit them when learning modals? Evidence from the acquisition literature suggests that they might. First, children seem to understand

[7] Modals also take an ordering source, which further constrains the set of worlds quantified over, and is responsible for meaning differences among root flavors.

[8] See also [38].

lexical aspectual distinctions very early in development ([33], [34]). Second, studies that have specifically examined modal flavor development in relation to complement type ([36], [37]) show that 3-year-olds may be sensitive to aspectual cues when interpreting modals.

5 Conclusions

The way speakers use modals makes it challenging to notice that some modals can express different flavors. Yet, children eventually pick up on this polysemy, and they may do so even earlier than the literature on the purported epistemic gap reports. We have proposed that one way children may pick up on modal flavor, and consequently modal polysemy, is by exploiting distributional properties that distinguish flavors. Speakers tend to use root modals with future TO, but not epistemic modals. If children expect correlations between modal flavor and TO, perhaps because they expect a constraint like the diversity condition to constrain modal meanings, they may exploit aspectual cues to discover modal meanings.

Modal polysemy seems to be, by and large, tied to functional modality. Might such syntactic bootstrapping overgenerate and trigger polysemous uses of monosemous, lexical modals? The constraints we have been discussing so far seem to be tied to notional modality, in ways that may transcend syntactic category: both functional and lexical modals with root meanings seem to be future-oriented, but not epistemic functional nor lexical modals [14]. Thus lexical modals that express root meanings perhaps rarely appear with non future TO, and won't lead learners astray. There may, however, be further constraints that uniquely apply to functional modality (e.g. scope interactions with tense, other modals, or quantifiers), which may require that learners be sensitive to the lexical status of their modals (see e.g., [37], [39]).

We have argued that a syntactic bootstrapping account where learners exploit aspectual properties to figure out modal meaning is plausible: the cues are clearly present in the input, the links between aspect and modal flavor are principled, and children seem to have the requisite understanding to exploit aspectual cues. We leave for future research whether children *do* in fact learn modal polysemy this way.

References

[1] van der Auwera, & Ammann (2011). Overlap between situational and epistemic modal marking. In Dryer & Haspelmath (eds.), *The WALS online,* 76. Max Planck Library.

[2] Clark (1987). The Principle of Contrast: A constraint on acquisition. In MacWhinney (ed.), *Proceedings of the 20th Carnegie Symposium on Cognition*, 1-33. Erlbaum.

[3] Coates (1988). The acquisition of the meanings of modality in children aged eight and twelve. *Journal of Child Language,* 15, 425-434.

[4] Papafragou (1998). The Acquisition of modality. *Mind and Language* 13 (3), 370-399.

[5] Kuczaj & Maratsos (1975). What children can say before they will. *Merrill-Palmer Quarterly of Behavior and Development* 21, 89–111.

[6] Wells (1979). Learning and using the auxiliary verb in English. In Lee (ed.), *Language Development,* 250-270. Croom Helm.
[7] Stephany (1979). Modality. In Fletcher & Garman (eds.) *Language Acquisition,* 375–400. CUP.
[8] Cournane (2015). Modal development. PhD. Thesis. UToronto.
[9] Sweetser (1990). From etymology to pragmatics. CUP.
[10] Veselinovic & Cournane (accepted). The grammatical source for missing epistemic meanings for modal verbs in child BCS. In *FASL 26.* Michigan Slavic Publications.
[11] O'Neill & Atance. (2000). The development of children's use of modals to express uncertainty. *First Language*, 20(58), 29–52.
[12] Condoravdi (2002). Temporal interpretation of modals. In Beaver, Casillas Martinez, Clark & Kaufmann (eds.) *The construction of meaning*, 59-88. CSLI.
[13] Werner (2006). Future and non-future modal sentences. *Natural Language Semantics,* 14, 235-255.
[14] Klecha (2016). Modality and embedded temporal operators. *Semantics and Pragmatics,* 9, 1-55.
[15] Rullmann & Matthewson. (in press). Towards A Theory of Modal-Temporal Interaction. To appear in *Language*.
[16] Theakston, Lieven, Pine, & Rowland (2001). The role of performance limitations in the acquisition of verb-argument structure. *Journal of Child Language*, 28, 127-152.
[17] MacWhinney (2000). The CHILDES project. Erlbaum.
[18] Kratzer (1981). The notional category of modality. In Eikmeyer & Rieser, (eds.), Words, Worlds, and Contexts, pages 38–74. De Gruyter.
[19] Cinque (1999). Adverbs and functional heads. OUP.
[20] Brennan (1993). Root and epistemic modal auxiliary verbs. Ph.D. Thesis. UMass Amherst.
[21] Stowell (2004). Tense and modals. In Guéron & Lecarme (eds.) The Syntax of Time, 495–537. MIT Press.
[22] Hacquard (2010). On the event relativity of modal auxiliaries. *Natural Language Semantics*, 18(1), 79-114.
[23] Kratzer (2012). Modals and conditionals: new and revised perspectives. OUP.
[24] Dowty (1979). Word meaning and Montague grammar. Kluwer.
[25] Gleitman (1990). The Structural Sources of Verb Meanings. *Language Acquisition,* 1, 3–55.
[26] Lidz (2006). Verb learning as a probe into children's grammars. In Hirsh-Pasek & Golinkoff (eds.) *When Action Meets Words*, 429-449. OUP.
[27] Gleitman, & Landau (1994). *Acquisition of the lexicon.* MIT Press.
[28] Gleitman, Cassidy, Nappa, Papafragou & Trueswell (2005). Hard words. *Language Learning and Development,* 1(1), 23–64.
[29] Papafragou, Cassidy & Gleitman (2007). When we think about thinking. *Cognition*, 105, 125–165.
[30] Hacquard & Lidz (to appear). Children's attitude problems. *Mind and Language.*
[33] Wagner (2001). Aspectual influences on early tense interpretation. *Journal of Child Language*, 28, 661–681.
[34] van Hout (2016). Lexical and Grammatical Aspect. In Lidz, Snyder, & Pater (eds.), *The Oxford Handbook of Developmental Linguistics.* OUP.
[36] Heizmann (2006). Acquisition of deontic and epistemic readings of must and müssen. In Heizmann (ed.) *UMass Occasional Papers in Linguistics 34*. GLSA.
[37] Hacquard (2013). On the grammatical category of modality. In Aloni, Franke & Roelofsen (eds.) *Proceedings of the 19th Amsterdam Colloquium,* 19–26.
[38] Thomas (2016). Circumstantial modality and the diversity condition. In Etxeberria, Fălăuş, Irurtzun, & Leferman (eds.), *Proceedings of Sinn und Bedeutung 18*, 433–450.
[39] Hacquard & Cournane (2016). Themes and variations in the expression of modality. *Proceedings of the 46th Annual Meeting of NELS*, 21-42.

Object Mass Nouns in Japanese

Kurt Erbach, Peter R. Sutton, Hana Filip, and Kathrin Byrdeck*

Heinrich-Heine University, Düsseldorf, Germany
erbach@hhu.de, peter.r.sutton@icloud.com, hana.filip@gmail.com, byrdeck@phil.hhu.de

Abstract

Classifier languages are commonly taken to have no grammaticized lexical mass/count distinction, but rather have this distinction encoded through the syntax and semantics of classifiers (e.g. [4], [5], [15], [17]). We contest this claim by drawing on data from Japanese. We provide novel empirical evidence showing that Japanese has quantifiers (e.g. *nan-byaku to iu* 'hundreds of') which directly select only for nouns denoting atomic entities (*onna no hito* 'woman') without requiring any classifier support. Moreover, the selectional restrictions of such quantifiers lead us to identify a class of object mass nouns in Japanese, i.e. nouns that have atomic entities in their denotation and yet are infelicitous in syntactic environments which are diagnostic of count nouns. This contradicts the prediction in [5] that object mass nouns should not exist in classifier languages. If Japanese has object mass nouns, then we should be ready to accept that Japanese nominal system is endowed with a grammatical mass/count distinction, and one which bears a certain resemblance to that which we find in number marking languages (e.g. English). We propose a novel semantic analysis of Japanese lexical nouns and classifiers, based on Sutton & Filip [21], a framework that unites notions of context in Rothstein [16] and Landman [12], and motivates the idea that counting contexts can remove overlap so that count nouns have disjoint counting bases while mass nouns do not.

1 Introduction

Japanese, a typical classifier language, is commonly taken to have no grammaticized lexical mass/count distinction, i.e. no lexical distinction between different kinds of nouns sensitive to countability that is reflected in the grammatical behavior of nouns. Instead, this sort of distinction is thought to be encoded through the syntax and semantics of classifiers (e.g. [4], [5], [15], [17]). However, we provide evidence that Japanese has quantifiers like (e.g. *nan-byaku to iu* 'hundreds of') that distinguish between mass and count nouns, whose denotation does not align with the semantic (ontological) non-atomic and atomic domains. This then motivates the existence of a group of nouns in Japanese with the two hallmark properties of object mass nouns: (i) they have atomic denotations, and (ii) are infelicitous in syntactic environments which are diagnostic of count nouns. Object mass nouns (alternatively *fake mass nouns*) are nouns such as *furniture* or *mail* in English, and are predicted to not exist in classifier languages [5]. Our results show that Japanese indeed has object mass nouns and *a forteriori* that the Japanese lexical nominal system has a mass/count distinction that is directly relevant to the grammar of Japanese. We do so by exploring the properties of Landman [12] and Sutton & Filip [21], we argue that the key factor underpinning the count/mass distinction is whether or not the entities that count as 'one' in the denotation of a noun (the counting base) overlap. Mass concepts have overlapping counting bases and count concepts have disjoint counting bases. Japanese quantifiers like *nan-byaku to iu* ('hundreds of'), we argue, can only compose with nouns that determine disjoint counting bases, without any classifier support. But this can be taken as evidence for the existence of bona fide count nouns in Japanese, and hence for countability having direct grammatical relevance for the Japanese grammar.

*This research is funded as part of DFG Collaborative Research Centre 991: The Structure of Representations in Language, Cognition, and Science. Our thanks to attendees of the 18th Szklarska Poreba Workshop, and especially to Yasutada Sudo and Eric McCready for valuable feedback. Thanks, too, to our consultants Kaori Fujita, Saki Kudo, Sebastian Steinfelder, Aiko Tendo, and Yuko Wagatsuma.

2 Background

Object mass nouns are of key importance in determining whether or not a language has a mass/count distinction, because they provide evidence for the mismatch between conceptual individuation, on the one hand, and grammatical mass behavior, on the other hand. We use the term *inherently individuable* to refer to entities that are *objects* as opposed to *substances* in the sense of Soja et al. [18]. Nouns with inherently individuable denotations can be count (e.g., *chair*, *cat*) or mass (e.g., *furniture*, *jewelry*). *Object mass nouns* are those nouns which have inherently individuable extensions, but that are nonetheless infelicitous in counting constructions (e.g. # *I bought two furnitures*). Chierchia's [5] explanation for object mass nouns is the *copy-cat effect*, according to which atomically stable nouns like *furniture* copy mass noun properties as a result of lexical choice. The theory of [5] predicts that object mass nouns are expected to be found in number marking languages like English, because their nouns are differentiated with respect to their denotations, and because lexical choice makes it simple to characterize a potential count noun as a mass noun. Object mass nouns cannot exist in classifier languages, according to [5], because all their nouns uniformly denote kinds, as they freely occur as bare nominal arguments and cannot directly compose with numericals (1):

(1) a. inu go-*(hiki)
dog five-CL$_{small.animal}$
'five dogs'[15, p. 73]

b. kagu itsu-*(tsu)
furniture five-CL$_{general}$
'five pieces of furniture'

c. yūbinbutsu go-*(bu)
mail five-CL$_{printed.material}$
'five pieces of mail'

d. mizu go-*(hon)
water five-CL$_{bottle}$
'five bottles of water'

The analysis of classifier languages in [5], and of the most influential to date, is couched in a compositional type-theoretic framework in which all nouns uniformly denote kinds ($\langle k \rangle$), and numericals are adjectival (of type $\langle \langle e,t \rangle, \langle e,t \rangle \rangle$); consequently, overt morphemes, namely classifiers of type $\langle k, \langle e,t \rangle \rangle$ must intervene between numericals and their nominal arguments.

There is, however, a growing body of work showing that a more nuanced view of the nominal system of classifier languages is warranted [1], [6], [9], [14], [19], [20]. For example, Inagaki & Barner [9] use comparison tasks in classifier-less 'more than' constructions, Japanese nouns like *kutsu* ('shoe') and *kagu* ('furniture') are compared according to cardinality of individuals, but substance nouns like *karashi* ('mustard') are judged according to volume. These 'more than' constructions were not only classifier-less but also lacked any other grammatical cues for individuation (i.e. the presence or absence of count syntax) that could have triggered a cardinality or volume comparison. Inagaki and Barner [9] take these results as evidence that some Japanese nouns encode the grammatical feature ±INDIVIDUATED even in the absence of classifiers or other count syntax.

In support of the stronger claim, that there are reflexes of the mass/count distinction in at least some classifier languages, Sudo [19], [20] argues that certain Japanese quantifiers differentially select for count nouns. For instance, *nan-byaku to iu* ('hundreds of') and *dono N mo* ('whichever' or 'every') are felicitous with count nouns (e.g. *hon* 'book') but infelicitous with mass nouns (e.g. *ase* 'sweat'). In [19], [20], this observation is taken to mean that there are nouns with count denotations in Japanese; i.e. the inherent individuation of extensions is directly encoded by Japanese nouns, rather than in count syntax via a classifier constructions.

This begs the question, however, why it is that count nouns can nonetheless not be directly modified by numerical expressions in Japanese. Sudo's [19] explanation of this (which mirrors one also found in Krifka [11]), is that the reason that numerical expressions in Japanese can only denote abstract objects of type $\langle n \rangle$. This differs from number marking languages, such as English, in which numerical expressions have a numerical determiner interpretation. On Sudo's analysis, classifiers denote functions which map entities of type $\langle n \rangle$ into expressions of

the adjectival modifier type $\langle s, \langle e, t \rangle \rangle$, which freely compose with common noun interpretations.

While Inagaki & Barner [9] show that Japanese nouns encode a feature ±INDIVIDUATED, Sudo [19], [20] makes the stronger claim that there are grammatical reflexes of a mass/count distinction in Japanese. However, if these reflexes were simply correlated with the atomic/non-atomic, or the individuated/non-individuated, distinction, then the analysis of Chierchia [3], [5] could be upheld by adding sensitivity to natural atomicity or individuation to the relevant parts of the grammar. In other words, a critic of Sudo could insist that classifier languages, such as Japanese, do not display a mass/count distinction in their nominal system, but merely mark the notional distinction between entities that are or are not inherently individuable.

One of the main contributions of this paper is to provide a means of resolving this dispute: *evidence for object mass nouns*. If the grammatical tests outlined by Sudo [19], [20] (such as felicitous combination with *nan-byaku to iu* 'hundreds of') can be shown to bisect the class of common nouns in a way that does not mirror the prelinguistic notional individuable/non-individuable divide, then we have evidence that the grammar encodes more than the mere notional distinction. In particular, if we find nouns with inherently individuable extensions that are infelicitous with e.g, *nan-byaku to iu* ('hundreds of'), we will have evidence that Japanese has grammatical reflexes a genuine lexical mass/count distinction. With this aim in mind we conducted an experiment designed to provide evidence for object mass nouns in Japanese.

3 Empirical Evidence

In English, object mass nouns, such as *furniture* have atomic denotations and yet are infelicitous with count quantifiers as for example *each* and *every* (2). For Japanese, it has been proposed that quantifiers such as *nan-byaku to iu* ('hundreds of') work similarly to *many*, in that it is felicitous with count nouns like *onna no hito* ('woman') but infelicitous with mass nouns like *yuki* ('snow') [19] as in (3).

(2) a. every dog b. *every furniture c. *every snow

(3) a. nan-byaku to iu onna.no.hito
what-hundred to say woman
'hundreds of women'

b. #nan-byaku to iu yuki
what-hundred to say snow
#'hundreds of snow'

3.1 Experimental Design

Building mainly on the observations about Japanese data in Sudo [20], we designed an acceptability judgment experiment in which we asked 49 native speakers (in an online survey on www.crowdworks.jp) to judge the acceptability of 120 sentences, including distractor sentences, on a five point Likert scale ranging from 1, *hen da* ('odd'), to 5, *yoi* ('good'). Each sentence contained a combination of the quantifier *nan-byaku to iu* ('hundreds of'), which does not require a classifier, with a noun. We tested 22 collective artifact nouns like *kagu* ('furniture') and *yūbinbutsu* ('mail') (6), alongside 11 nouns denoting discrete entities/individuals (e.g. *onna no hito* 'woman' in (5)) and 11 nouns denoting undifferentiated stuff like *yuki* ('snow') in (4). Sentences with an average acceptability rating higher than the neutral rating 3 were categorized as felicitous, whereas sentences with an average rating lower than 3 were categorized as infelicitous and marked accordingly in (4)-(6).

(4) kinō yuki ga fu-tta. #nan-byaku to iu yuki wa mō toke-te
yesterday snow NOM fall-PST; #what-hundred to say snow NOM already melt-TE
shima-tta
finish-PST
'It snowed yesterday. #Hundreds of snow melted already.'

(5) toranpu-shi ga daitoryō ni na-tta ato, nan-byaku to iu
Trump-president NOM president ACC become-PST after; what-hundred to say
onna.no.hito ga washinton de neriarui-ta
woman NOM Washington LOC march-PST
'After Trump became president, hundreds of women marched in Washington DC.'

(6) kono yūmei-na aidorugurūpu wa fanretā ga aoku-te pinku no fūtō dake
this famous-ADV band TOP fanletter NOM blue-TE pink GEN envelope only
de mora-tte iru. #senshū mo nan-byaku to iu yūbinbutsu o mora-tte
with become-TE PRG; #lastweek too what-hundred to say mail ACC get-TE
i-ta
PRG-PST
'This famous band gets fan letters exclusively in pink and blue envelopes. Last week they got #hundreds of mail.'

3.2 Results

The main results are summarized in Figure 1. Across participants, judgments were found to be consistent using the Friedman test [8], meaning there was little variance in judgment per test item. Nouns denoting discrete entities (e.g. *onna no hito* 'woman') were judged to be felicitous with *nan-byaku to iu* ('hundreds of'), with the average judgment of 3.92. Nouns like *yuki* ('snow') denoting undifferentiated stuff had an average judgment of 2.08, and were infelicitous with *nan-byaku to iu* ('hundreds of'). The collective artifact denoting noun *yūbinbutsu* ('mail') is also infelicitous with *nan-byaku to iu* ('hundreds of'), receiving an average judgment of 2.25.

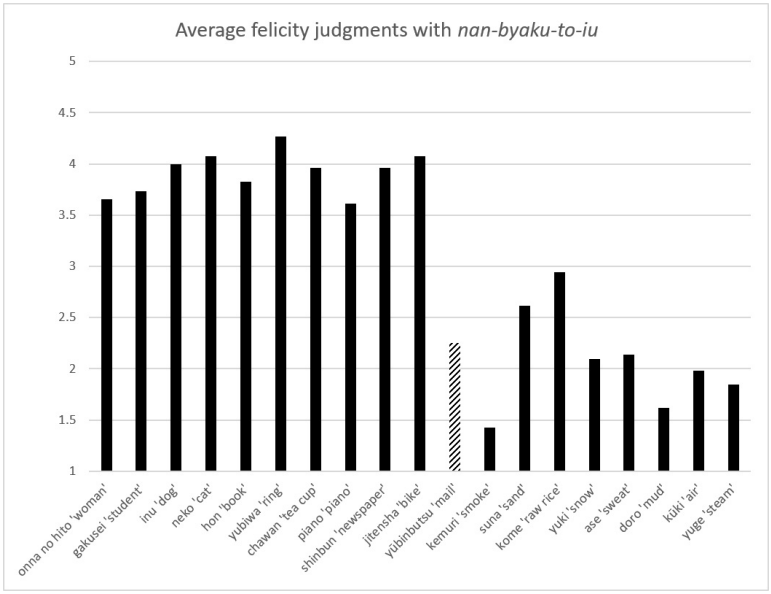


Figure 1: Bi-partite split of Japanese nouns based on compatibility with *nan-byaku to iu* ('hundreds of')

3.3 Discussion

The two competing hypotheses regarding the selectional restrictions of *nan-byaku to iu* ('hundreds of') are: (i) *nan-byaku to iu* ('hundreds of') is a suitable test of whether the extension of its argument noun has inherently individuable structure; (ii) *nan-byaku to iu* ('hundreds of') is a suitable test of whether its argument noun is count. If hypothesis (i) were correct, we would expect to see low judgement scores for all nouns that lack inherently individuable extensions and high scores for all nouns that have inherently individuable extensions. Evidence against hypothesis (i) and in favor of hypothesis (ii) would be for felicity scores with *nan-byaku to iu* ('hundreds of') to form a partition that does not mirror the individuable/non-individuable divide.

The results for *yūbinbutsu* ('mail') provide exactly the evidence we were looking for in support of hypothesis (ii). Although *nan-byaku to iu* ('hundreds of') is infelicitous with all nouns that denote substances (which lack an inherently individuable structure), *nan-byaku to iu* ('hundreds of') is not felicitous with all nouns that denote objects (which have an inherently individuable structure), namely *yūbinbutsu* ('mail'). In the absence of an alternative explanation for this pattern, we have good reason to conclude that Japanese has grammatical reflexes of the lexicalized mass/count distinction, and what is more, it also has object mass nouns. Both of these conclusions conflict with the common view of the nominal system in classifier languages, as, for instance, implemented in Chierchia's [5] recent analyses of the nominal semantics for classifier languages.

One possible counterargument to our conclusions, however, would be that *yūbinbutsu* ('mail') actually does not denote entities with an inherently individuable structure (at least in the way that Japanese speakers perceive of them). To reject this counterargument, we have begun to test native speaker judgements using the 'more than' test [9]. If a noun denotes entities with an inherently individuable structure, then there should be a felicitous *cardinality comparison* reading available for questions with 'more than'. If a noun denotes entities which lack an inherently individuable structure, then there should only be a felicitous *measure comparison* reading available for questions with 'more than'. To determine which of these options applies to *yūbinbutsu* ('mail'), we presented native speakers with sentences in which a measure or cardinality comparison is possible between two groups of items. Each sentence used one of our test nouns, and each had a group larger in volume but smaller in cardinality—e.g.s (7)-(9).

(7) Yuma wa futa-tsu no fūtō o uketo-tta. Hito-tsu wa atarashī shigoto
Yuma TOP 2-CL GEN envelopes ACC receive-PST. 1-CL TOP new work
no keiyaku de, mō-hito-tu wa apāto no keiyaku da. Satomi wa
GEN contract and, another-1-CL TOP apartment GEN contract COP. Satomi TOP
itsu-tsu no chīsai fūtō o uketo-tta. Doremo tomodachi kara no chīsai
5-CL GEN small envelope ACC receive-PST. Both friend from GEN small
tegami o fukun-de iru.
letter ACC contain-TE IRU.
'Yuma received two large envelopes, one with her new work contract and one with her apartment contract. Satomi got five small envelopes, each containing a short letter from a friend.'

(8) Mai wa yot-tsu no ōkī koshikake o ka-tta. Hiroaki wa itsu-tsu no
Mai TOP 4-CL GEN big armchair ACC buy-PST. Hiroaki TOP 5-CL GEN
kodomo-yō no chīsai isu o ka-tta.
child-use GEN small chair ACC buy-PST.
'Mei bought four large arm chairs. Hiroaki bought five small chairs for children.'

(9) Toma wa hito-tsu no ōkī yukidaruma o tsuku-tta. Mizuki wa itsu-tsu no
Toma TOP 1-CL GEN big snowman ACC make-PST. Mizuki TOP 5-CL GEN
yuki no tama o tsuku-tta.
snow GEN ball ACC make-PST.
'Toma made a big snowman. Isuki made five small snowballs.'

Following the presentation of each scenario, we asked the speakers to judge who has more *yūbinbutsu* ('mail')*isu* ('chair') and *yuki* ('snow'), respectively. In our pretest, *yūbinbutsu* ('mail') and *isu* ('chair') were judged according to cardinality comparison, while *yuki* ('snow') was judged according to volume. This is evidence that the extension of *yūbinbutsu* ('mail') has an inherently individuable structure.

In sum, the above data leads us to the conclusion that the Japanese nominal system does not only distinguish the notional individuable/non-individuable divide, but, in fact, has grammatical reflexes of the mass/count distinction, as attested by the presence of nouns which denote entities with an individuable structure, but nonetheless pattern, when combined with *nan-byaku to iu* ('hundreds of'), with substance denoting nouns. In other words, for Japanese we found evidence for the existence of object mass nouns, namely, *yūbinbutsu* ('mail').

4 Analysis

Our quantification and quantity judgment data respectively show clear grammatical and notional differences between Japanese nouns. To account for these grammaticized lexical differences in Japanese, we build on Sutton & Filip [21], who argue that the grammaticized lexical mass/count distinction is grounded in the (non-)resolution of overlap (also see [12]). To their model of lexical entries, we add a field for presuppositions (or, more neutrally, preconditions) for composition. We use presuppositions in two main ways: (i) in the entries for sortal classifiers, they capture selectional restrictions on the nouns with which they combine (e.g., that the argument noun must denote printed items); (ii) in the entries for sortal classifiers and count quantifiers, they require counting bases of argument nouns to be disjoint. In Section 4.1, outline an account of the semantics of the mass/count distinction in English (based on [21] and [13]). In Section 4.2, we extend this account to cover the Japanese data by providing an analysis of Japanese numerical expressions, classifiers, and *nan-byaku-to-iu* ('hundreds of').

4.1 Counting in context

Sutton & Filip [21] provide a cross-linguistic analysis of collective artifact nouns, such as *furniture* and *kitchenware*, in English. The puzzle they address is why collective artifact nouns stubbornly resist count-to-mass coercion when directly modified with a numerical expression (*# three furnitures/kitchenwares)*. Their solution is based on exploiting two types of counting contexts: *specific counting contexts*, which remove overlap in counting bases (the set of entities for counting); and *null counting contexts*, which allow overlap in counting bases.

Recent semantic analyses of the count/mass distinction [12, 13, 21] advocate representing the lexical entries of concrete nouns using ordered pairs. For example Landman [13] represents CN entries as $\langle body, base \rangle$, a pair of *base*, the counting base set, and *body*, a subset of the upward closure of *base* under sum. Following Krifka [10], Sutton & Filip analyze the lexical entries of nouns as including qualitative and quantitative criteria of application in the lexical entries of nouns. They are presented as ordered pairs, $\langle P, counting\_base(P) \rangle$. P is a property for the qualitative criteria of applying the noun concept. $counting\_base(P)$ specifies the quantitative criteria for applying the noun concept, which, crucially, includes information regarding: (a) whether or not the extension is inherently individuable; and (b) whether or not potentially countable entities are conceptualised in terms of a disjoint individuation schema (formalised in terms of counting contexts). Counting goes wrong when the counting base is an overlapping

set. Grammatical counting is possible when the counting base is a disjoint set.

Here we combine some elements of Landman's account (distinguishing our *body* from *base*) and some of Sutton & Filip's (inclusion of an individuation function interpreted at a counting context). Furthermore, following Filip & Sutton [7], we introduce a third projection to record preconditions and/or presuppositions relating to e.g., selectional restrictions, so CN entries have the form $\langle extension,\ c\_base,\ presup\rangle$.

The components *extension* and *c_base* are formed from up to three ingredients: A predicate, $P_{\langle e,t\rangle}$, an $\mathbf{IND}_{\langle\langle e,t\rangle,\langle e,t\rangle\rangle}$ function, and a counting context $c_{\langle\langle e,t\rangle,\langle e,t\rangle\rangle}$. For example $\mathbf{IND}(CAT)$ denotes the disjoint set of single cats. However, the **IND**-set is not always disjoint for other predicates. For example, the set of things that count as one for collective artifact nouns overlaps [12, 13, 21] e.g., a nest of tables ($a_1 \sqcup a_2 \sqcup a_3$), and the individual tables in the nest (a_1, a_2, a_3) each count as one with respect to *furniture*): $\{a_1, a_2, a_3, a_1 \sqcup a_2 \sqcup a_3\} \subset \mathbf{IND}(FURN)$.

Further 'perspectives' on **I**ND-sets are represented with counting contexts (of type $\langle\langle e,t\rangle,\langle e,t\rangle\rangle$) which come in two varieties:

Specific counting contexts $c_{i\geq 1}$: map sets onto maximally disjoint subsets. Intuitively, the specific counting contexts represent the imposition of a disjoint individuation schema.

The null counting context $c_{i\geq 1}$: is the identity function. It does not remove overlap if present: $\forall P \forall x[\mathrm{IND}(P)(x) \leftrightarrow c_0(\mathrm{IND}(P))(x)]$

Overlapping counting bases give rise to mass predicates, since grammatical counting requires disjointness. Therefore, evaluated at a specific counting context, the set denoted by $c_{i\geq 1}(\mathbf{IND}(FURN))$ is disjoint and disjoint counting bases mean grammatical countability. Evaluated at the null counting context c_0, the set denoted by $c_0(\mathbf{IND}(FURN))$ is overlapping and overlapping counting bases mean grammatical non-countability. Notice, however, that if an **IND**-set is anyway disjoint, there is no difference whether it is evaluated at a specific counting context or at the null counting context. Sutton & Filip [21] argue that this accounts for cross-linguistic variation in mass/count lexicalization patterns for collective artifact nouns. Whether or not a lexical entry indexes the **IND**-set to the null counting context or to a specific counting context is essentially a matter of lexical 'choice' (a parameter set language by language and noun by noun). This explains why nouns such as *cat*, and its cross-linguistic counterparts are all lexicalized as count ($\forall c_i[c_0(\mathbf{IND}(CAT)) = c_i(\mathbf{IND}(CAT))]$). It also explains why nouns which denote inherently individuable entities, but for which the **IND**-set of entities that count as one overlap can be lexicalized as either count or mass cross- and intra-linguistically. For example, we have an account for why we find the count noun *meubel* ('(piece of) furniture', Dutch) as well as the mass nouns *furniture* and *meubilair* ('furniture', Dutch).

Sutton & Filip also argue that predicates for substances and objects are semantically distinguished, which is supported by the ability of pre-linguistic infants to distinguish substances from objects [18]. Formally, this translates as there being no **IND** function in the lexical entries for substance denoting nouns (nouns which denote stuff that lacks an inherently individuable structure). Importantly however, the distinction between substances and objects does not perfectly mirror the mass/count distinction, as seen in the behavior of nouns like *furniture* which have objects in their denotation, yet grammatically pattern with nouns that denote substances, liquids, and gases. (For an explanation of why substance denoting nouns are almost always, but not universally lexicalized as mass, see Sutton & Filip [22].) Examples of a range of lexical entries are given in (1a–1f):

$$[\![cat]\!]^{c_i} = \lambda x.\langle c_i(\mathbf{IND}(\mathrm{CAT}))(x),\ \lambda y.c_i(\mathbf{IND}(\mathrm{CAT}))(y),\ \varnothing\rangle \tag{1a}$$

$$[\![cats]\!]^{c_i} = \lambda x.\langle {}^*c_i(\mathbf{IND}(\mathrm{CAT}))(x),\ \lambda y.c_i(\mathbf{IND}(\mathrm{CAT}))(y),\ \varnothing\rangle \tag{1b}$$

$$[\![meubel]\!]^{c_i} = \lambda x.\langle c_i(\mathbf{IND}(\mathrm{FURN}))(x),\ \lambda y.c_i(\mathbf{IND}(\mathrm{FURN}))(y),\ \varnothing\rangle \tag{1c}$$

$$[\![meubels]\!]^{c_i} = \lambda x.\langle {}^*c_i(\mathbf{IND}(\text{FURN}))(x),\ \lambda y.c_i(\mathbf{IND}(\text{FURN}))(y),\ \varnothing\rangle \quad (1d)$$

$$[\![furniture]\!]^{c_i} = \lambda x.\langle {}^*c_0(\mathbf{IND}(\text{FURN}))(x),\ \lambda y.c_0(\mathbf{IND}(\text{FURN}))(y),\ \varnothing\rangle \quad (1e)$$

$$[\![mud]\!]^{c_i} = \lambda x.\langle {}^*c_0(\text{MUD})(x),\ \lambda y.c_0(\text{MUD})(y),\ \varnothing\rangle \quad (1f)$$

Each entry contains *extension* (the truth conditions for applying the noun), *c_base* (the individuation schema for the noun concept), and *presup* (a slot which can specify extra lexical or compositional information or restrictions). The singular nouns *cat* and *meubel* ('(item of) furniture', Dutch) in (1a) and (1c) are interpreted at the specific counting context of utterance c_i. This removes any overlap in the counting bases. Their application conditions and individuation schemas express the same properties (the sets of single cats/items of furniture) and we get the grammatical count nouns *cat* and *meubel* ('(item of) furniture', Dutch). The plural forms ((1b) and (1d)) require the extensions to be single cats/items of furniture or sums thereof. The mass nouns *furniture* and *mud* in (1e) and (1f) are interpreted relative to the null counting context c_0. This allows overlap in the counting bases (i.e. different overlapping partitions of mud-stuff or different overlapping partitions of furniture into items), and so we get the grammatically mass nouns *mud* and *furniture*.

In short, the only difference in the entries for the plural count noun *meubels* and the mass noun *furniture* is whether the counting base is interpreted at c_0 or c_i. Interpretation at the null or at a specific counting context is essentially a matter of lexical choice. Hence, we expect both count and mass terms, cross-linguistically, to express this concept.

4.2 Nominal semantics in Japanese

Lexical entries for common nouns. On our analysis, lexically simple Japanese nouns have lexical entries that closely match those in number marking languages. Count nouns like *isu* ('chair') are interpreted at a specific counting context that specifies disjoint counting base 2b. Object mass nouns like *yūbinbutsu* ('mail', 2c) and *yuki* ('snow', 2a) have entries saturated with the null counting context c_0, but *yuki* ('snow', 2a), as a substance denoting noun is interpreted without the **IND**-function. The one difference between e.g., Japanese and English is that, since Japanese has a highly restricted (and even then, optional) use of plural morphology, lexically simple Japanese nouns have number neutral extensions (that include entities and sums thereof).

$$[\![yuki]\!]^{c_i} = \lambda x.\langle {}^*c_0(\text{SNOW})(x),\ \lambda y.c_0(\text{SNOW})(y),\ \varnothing\rangle \quad (2a)$$

$$[\![isu]\!]^{c_i} = \lambda x.\langle {}^*c_i(\mathbf{IND}(\text{CHAIR}))(x),\ \lambda y.c_i(\mathbf{IND}(\text{CHAIR}))(y),\ \varnothing\rangle \quad (2b)$$

$$[\![yūbinbutsu]\!]^{c_i} = \lambda x.\langle {}^*c_0(\mathbf{IND}(\text{MAIL}))(x),\ \lambda y.c_0(\mathbf{IND}(\text{MAIL}))(y),\ \varnothing\rangle \quad (2c)$$

Counting with classifiers in context. Both count nouns and object mass nouns can be combined with a numerical expression when there is an intervening sortal classifier. Following Krifka [11], we assume that numericals (e.g. 3a) denote numbers of type $\langle n\rangle$. Key to our analysis are four functions for Japanese, sortal classifiers, e.g. *bu* ('printed item') in 3b: (1) they map type n expressions to expressions of the type for numerical determiners; (2) they also ensure that the counting base predicate provided by the argument noun is evaluated at the counting context of utterance. For example, if the argument noun is saturated with the null counting context, then the equivalence in 3c ensures that overlap is removed in the resulting counting base predicate; (3) they add a presupposition that the counting base is disjoint (so as to be fit for counting); (4) they add a presupposition that the argument predicate is of some restricted sort. For example, for *bu* ('printed item'), it is the presupposition that the argument predicate denotes a subset of PRINTED.ITEM (this also acts to filter out combination

with substance denoting nouns).[1]

$$[\![san]\!]^{c_i} = 3 \tag{3a}$$

$$[\![bu]\!] = \lambda n.\lambda P.\lambda c.\lambda x.\langle \pi_1(P(x)),\ \mu_{card}(x, \lambda y.c(\pi_2(P(x)))(y)) = n, \tag{3b}$$
$$\text{DISJ}(\lambda y.c(\pi_2(P))(x)(y))\ \wedge\ \lambda x.\pi_1(P(x)) \subseteq \text{PRINTED.ITEM}\rangle$$

$$\forall P.\forall c.[c(c_0(P)) \longleftrightarrow c(P)] \tag{3c}$$

$$[\![y\bar{u}binbutsu\ san\text{-}bu]\!]^{c_i} = \lambda x.\langle {}^{*}c_0\mathbf{IND}(\text{MAIL}(x),\ \mu_{card}(x, \lambda y.c_i(\mathbf{IND}(\text{MAIL})(y)) = 3, \tag{3d}$$
$$\text{DISJ}(\lambda y.c_i(\mathbf{IND}(\text{MAIL})(y))\ \wedge$$
$${}^{*}c_0\mathbf{IND}(\text{MAIL}) \subseteq \text{PRINTED.ITEM}\rangle$$

The result of combination, expressed in 3d, is the set of items of mail that have cardinality 3 at the counting context of utterance under the presupposition that the set of single items is disjoint and that mail is a type of printed material.

The quantifier *nan-byaku-to-iu* ('hundreds of') has, on our analysis, a semantics that closely resembles that of a numerical combined with a sortal classifier. The key difference is that it does not introduce a new context variable (intuitively, it does not provide information for selecting a disjoint individuation schema). Other differences are that the cardinality it specifies is underspecified (which we represent with the context-determined type n variable n_c), and *nan-byaku-to-iu* ('hundreds of') does not place extra restrictions (e.g., printed item) on the argument noun.

$$[\![nan\text{-}byaku\text{-}to\text{-}iu]\!] = \lambda P.\lambda x.\langle \pi_1(P(x)),\ \mu_{card}(x,\ \pi_2(P(x))) \geq n_c,\ \text{DISJ}(\pi_2(P(x)))\rangle \tag{4a}$$

$$[\![nan\text{-}byaku\text{-}to\text{-}iu\ isu]\!] = \lambda x.\langle \text{CHAIR}(x),\ \mu_{card}(x,\ \lambda y.\ c_i(\mathbf{IND}(\text{CHAIR})(y)) \geq n_c, \tag{4b}$$
$$\text{DISJ}(\lambda y.\ c_i(\mathbf{IND}(\text{CHAIR})(y))\rangle$$

$$[\![nan\text{-}byaku\text{-}to\text{-}iu\ y\bar{u}binbutsu]\!] = \lambda x.\langle \text{MAIL}(x),\ \mu_{card}(x,\ \lambda y.\ c_i(\mathbf{IND}(\text{MAIL})(y)) \geq n_c, \tag{4c}$$
$$\text{DISJ}(\lambda y.\ c_0(\mathbf{IND}(\text{MAIL})(y))\rangle$$
$$\Rightarrow \text{FALSE PRECONDITION!}$$

This simple difference is enough to capture the selectional restrictions of *nan-byaku-to-iu* ('hundreds of') since it predicts that *nan-byaku-to-iu* ('hundreds of') will only straightforwardly felicitously combine with count nouns. In 4b, $\lambda y.\ c_i(\mathbf{IND}(\text{CHAIR})(y)$ is a disjoint set, so *isu* ('chair') is count and *nan-byaku-to-iu isu* is felicitous. In 4c, $\lambda y.\ c_0(\mathbf{IND}(\text{MAIL})(y)$ is not disjoint, so *yūbinbutsu* is mass and *nan-byaku-to-iu yūbinbutsu* is infelicitous.

5 Conclusion

Our novel empirical evidence confirms that the Japanese quantifier *nan-byaku to iu* 'hundreds of' is a suitable diagnostic test for the count status of Japanese nouns. Moreover, and more importantly, we show that Japanese has object mass nouns, contrary to the prediction in [5] that they should not exist in classifier languages. This has not been shown in any previous work on classifier languages, to the best of our knowledge. Based on our findings for Japanese, we reject the common view that the mass/count distinction in *all* classifier languages is solely reflected in the syntax and semantics of their classifier systems, advocated for in [4], [5] or [15] among others. In Japanese, we find direct grammatical reflexes of the grammaticized

[1] Lexical entries for classifiers make use of product types (e.g.[2]). For example, an expression $\langle X_a, Y_b, Z_c\rangle$ is of type $\langle a \times b \times c\rangle$. We also use projection functions π_1 and π_2 such that $\pi_1(\langle X_a, Y_b, Z_c\rangle)=X_a$ and $\pi_2(\langle X_a, Y_b, Z_c\rangle)=Y_b$.

lexical mass/count distinction, as we argue. If there are classifier languages like Japanese that have a grammatical mass/count distinction in the lexicon, then the nominal system of such classifier languages are typologically closer to the nominal systems in languages with a bona fide lexical mass/count distinction, like English, than has previously been assumed. This conclusion requires a novel formal analysis of Japanese nouns, numericals, classifiers, and quantifiers, which we have provided based on [21].

References

[1] Alan Bale and David Barner. Semantic triggers, linguistic variation and the mass-count distinction. In *Count and mass across languages*, pages 238–260. Oxford University Press, 2012.

[2] Bob Carpenter. *Type-logical semantics*. MIT press, 1997.

[3] Gennaro Chierchia. Plurality of mass nouns and the notion of "semantic parameter". In *Events and Grammar: Studies in Linguistics and Philosophy Vol. 7*, pages 53–103. Kluwer, 1998.

[4] Gennaro Chierchia. Reference to kinds across language. *Natural language semantics*, 6(4):339–405, 1998.

[5] Gennaro Chierchia. Mass nouns, vagueness and semantic variation. *Synthese*, 174(1):99–149, 2010.

[6] Jenny Doetjes. Count/mass distinctions across languages. In Claudia Maienborn, Klaus von Heusinger, and Paul Portner, editors, *Semantics: An International Handbook of Meaning*, volume 3, pages 2559–2580. Walter de Gruyter, 2012.

[7] Hana Filip and Peter Sutton. Singular count nps in measure constructions. *Semantics and Linguistic Theory*, 27(0), forthcoming.

[8] Milton Friedman. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the american statistical association*, 32(200):675–701, 1937.

[9] Shunji Inagaki and David Barner. Countability in absence of count nouns: Evidence from japanese quantity judgments. *Studies in language sciences*, 8:111–125, 2009.

[10] Manfred Krifka. Nominal reference, temporal constitution and quantification in event semantics. In Renate Bartsch, J. F. A. K. van Benthem, and P. van Emde Boas, editors, *Semantics and Contextual Expression*, pages 75–115. Foris Publications, 1989.

[11] Manfred Krifka. Common nouns: A contrastive analysis of English and Chinese. In G.N. Carlson and F. J. Pelletier, editors, *The Generic Book*, pages 398–411. Chicago University Press, 1995.

[12] Fred Landman. Count nouns-mass nouns, neat nouns-mess nouns. *Baltic International Yearbook of Cognition, Logic and Communication*, 6(1):12, 2011.

[13] Fred Landman. Iceberg semantics for count nouns and mass nouns: The evidence from portions. *The Baltic International Yearbook of Cognition Logic and Communication*, 11, 2016.

[14] Peggy Li, Yarrow Dunham, and Susan Carey. Of substance: The nature of language effects on entity construal. *Cognitive psychology*, 58(4):487–524, 2009.

[15] Keiko Muromatsu. Classifiers and the count/mass distinction. *Functional structure (s), form and interpretation: Perspectives from East Asian languages*, pages 65–96, 2003.

[16] Susan Rothstein. Counting and the mass/count distinction. *Journal of semantics*, 27(3):343–397, 2010.

[17] Susan Rothstein. *Semantics for Counting and Measuring*. Cambridge University Press, 2017.

[18] Nancy N Soja, Susan Carey, and Elizabeth S Spelke. Ontological categories guide young children's inductions of word meaning: Object terms and substance terms. *Cognition*, 38(2):179–211, 1991.

[19] Yasutada Sudo. The semantic role of classifiers in japanese. *Baltic International Yearbook of Cognition, Logic and Communication*, 11(1):10, 2016.

[20] Yasutada Sudo. Countable nouns in japanese. *Proceedings of WAFL*, 11(1):11, 2017.

[21] Peter Sutton and Hana Filip. Counting in context: count/mass variation and restrictions on coercion in collective artifact nouns. *Semantics and Linguistic Theory*, 26(0):350–370, 2016.

[22] Peter Sutton and Hana Filip. Individuation, reliability, and the mass/count distinction. *Journal of Language Modelling*, 5(2):303–356, 2017.

Movement and alternatives don't mix: Evidence from Japanese*

Michael Yoshitaka Erlewine[1] and Hadas Kotek[2]

[1] National University of Singapore mitcho@nus.edu.sg
[2] New York University hadas.kotek@nyu.edu

Abstract

Certain quantificational elements ("interveners") have long been known to disrupt the interpretation of *wh*-in-situ (Hoji 1985 and many others), but the correct description of the set of interveners and the nature of intervention effects have been the subject of continued debate. In Erlewine and Kotek (2017), we offer a new generalization concerning the nature of intervener-hood in Japanese: A quantifier acts as an intervener if and only if it is scope-rigid. We argue that this generalization is explained by — and in turn supports — Kotek's (2017) account of intervention effects as reflecting a logical incompatibility between Predicate Abstraction and the computation of Rooth-Hamblin alternatives. In this paper we provide additional evidence in support of the above generalization, and test several of its predictions.

1 Intervention and intervener-hood

This paper concerns the proper characterization of so-called *intervention effects* in *wh*-questions and the characterization of interveners in Japanese. Intervention effects refer to the inability of certain quantificational elements to precede an in-situ *wh*-phrase, in a c-commanding position at surface structure. For example, Hoji (1985) observes that a *wh*-MO universal quantifier cannot precede a *wh* object in canonical in-situ position (1).[1]

(1) **Intervention with universal *wh*-mo:** (Hoji 1985:270)

?? **Da're-mo**-ga *nani*-o kai-mashi-ta-ka?
who-MO-NOM what-ACC buy-POLITE-PAST-Q
Intended: 'What did everyone buy?'

However, not all quantificational elements trigger intervention. For example, as noted by Tomioka (2007:1574), the universal quantifier *subete-no*-NP 'all NP' in the same configuration as in (1) does not lead to ungrammaticality:

(2) **Universal *subete* 'all' does not cause such intervention:**

✓ [**Subete**-no hito]-ga *nani*-o kai-mashi-ta-ka?
all-GEN person-NOM what-ACC buy-POLITE-PAST-Q
'What did everyone buy?'

*For comments and questions on this work, we thank participants of the NYU seminar on *wh*-constructions cross-linguistically and the NUS syntax/semantics reading group, as well as audiences at LENLS 2017, Stony Brook University, and the University of Pennsylvania. For discussion of judgments, we thank Minako Erlewine, Hiroki Nomoto, Yohei Oseki, and Yosuke Sato. Errors are each other's.

[1] Throughout the paper, interrogative *wh* are in *italics* and quantifiers of interest (potential interveners) — as well as sentential negation below — are in **bold**.

Even without changing the choice of intervener, Hoji (1985) notes that scrambling the *wh* in (1) above the quantifier also yields a grammatical question, as in (3).

(3) **Intervention is avoided by scrambling the intervener**

✓ *Nani*-o **da're-mo**-ga ___ kai-mashi-ta-ka?
what-ACC who-MO-NOM buy-POLITE-PAST-Q
'What did everyone buy?'

What makes the *wh*-MO universal quantifier (1) an intervener but not the *subete* universal quantifier (2)? More generally: What is the proper characterization of the set of interveners, and what is the nature of intervention? Previous work has tied intervention — and therefore the set of intervening elements — to the semantics of focus (Kim 2002, Beck 2006, Beck and Kim 2006), quantification (Beck 1996), topichood (Grohmann 2006), prosody (Tomioka 2007), (anti-)additivity (Mayr 2014), and semantic type-mismatch (Li and Law 2016).

Against this backdrop, we showed in Erlewine and Kotek 2017 that intervener-hood tracks scope-rigidity in Japanese. For example, even though the two universal quantifiers in (1–2) may have the same denotation as a universal quantifier, they differ in their scope-rigidity with respect to negation:

(4) ***wh-mo* universal quantifier is scope-rigid; *subete* is not:**

a. **Da're**-o-**mo** tsukamae-**nak**-atta.
who-ACC-MO catch-NEG-PAST
'*pro* did not catch anyone.' ✓every > not, *not > every

b. [**Subete**-no mondai]-o toka-**nak**-atta. (Mogi 2000:59)
all-GEN problem-ACC solve-NEG-PAST
'*pro* did not solve every problem.' ✓every > not, ✓not > every

Shibata (2015a) reports a similar correlation: *ka*-disjunction is scope-rigid with respect to negation whereas *naishi*-disjunction is not (5), and this correlates with intervener-hood (6).[2]

(5) ***ka*-disjunction is scope-rigid; *naishi* is not:**

a. [Taro-**ka** Jiro]-ga ko-**nak**-atta. (Shibata 2015a:23)
Taro-or Jiro-NOM come-NEG-PAST
'Taro or Jiro didn't come.' ✓or > not, *not > or

b. [Taro-**naishi** Jiro]-ga ko-**nak**-atta. (Shibata 2015a:96)
Taro-or Jiro-NOM come-NEG-PAST
'Taro or Jiro didn't come.' ✓or > not, ✓not > or

(6) ***ka* is an intervener; *naishi* is not:**

a. ??? [Taro-**ka** Jiro]-ga *nani*-o yon-da-no? (Hoji 1985:264)
Taro-or Jiro-NOM *what*-ACC read-PAST-Q

b. ✓ [Taro-**naishi** Jiro]-ga *nani*-o yon-da-no?
Taro-or Jiro-NOM what-ACC read-PAST-Q
'*What* did [Taro or Jiro] read?' (Shibata 2015a:98)

[2] We note that many speakers, including the first author, do not have clear judgments for *naishi* or feel that *naishi* simply patterns together with *ka* in (5–6). The judgments in (5–6) are those reported by Shibata. There seem to also be speakers who allow the 'not > or' reading of *ka* in (5) and for whom *ka* is not an intervener; Daisuke Bekki (p.c.) notes that he is such a speaker. What is important here is simply that there is a correlation between scope-rigidity and intervener-hood.

Erlewine and Kotek 2017 shows that this correlation generalizes across a variety of quantificational elements in Japanese, as summarized in (7). Here, "Scope-rigid" (◯) indicates that the given quantifier takes obligatory wide scope with respect to negation, whereas non-"scope-rigid" (×) quantifiers exhibit scope ambiguities with respect to negation. The nature of such scope ambiguities will be discussed in section 2.2 below.

(7) **Summary of Japanese data from Erlewine and Kotek 2017:**

	disjunction		universal		NPI only	NPI	modified
	ka	*naishi*	*wh-mo*	*subete*	*-shika*	*wh-mo*	numerals
scope-rigid?	◯ (5a)	× (5b)	◯ (4a)	× (4b)	◯ (K:228)	◯[3]	× (S:66)
intervener?	◯ (6a)	× (6b)	◯ (1)	× (2)	◯ (DT:134)	◯ (EK:4)	× (EK:5)

	indefinite	also	even	only	
	wh-ka	*-mo*	*-sae*	-P-*dake*	*-dake*-P
scope-rigid?	◯ (S:72)	◯ (M:59)	◯ (M:59)	◯ (F:12)	× (F:12)
intervener?	◯ (HH:269)	◯ (NH:119; Y:30)	◯ (Y:30)	◯ (EK:6)	× (EK:6)

Abbreviations: "*X*:*pp*" = *X* page *pp*; F = Futagi 2004; K = Kataoka 2006; M = Mogi 2000; HH = Hoji 1985; NH = Hasegawa 1995; S = Shibata 2015a; DT = Takahashi 1990; ST = Tomioka 2007; Y = Yanagida 1996; EK = Erlewine and Kotek 2017

Based on this evidence, we offered the following generalization in Erlewine and Kotek (2017):

(8) **Generalization: Intervention correlates with scope-taking**
Scope-rigid quantifiers above an in-situ *wh* cause intervention. Quantifiers that allow scope ambiguities — i.e., those that allow reconstruction below *wh* — do not.

We propose that the generalization in (8) can be derived based on Kotek's (2017) account for intervention effects, as a corollary of a logical incompatibility between Predicate Abstraction and Rooth-Hamblin alternative computation (see e.g. Shan 2004, Novel and Romero 2009, Ciardelli, Roelofsen, and Theiler 2017, Charlow 2017). In section 2, we briefly present the Kotek 2017 theory for intervention and then explain how this derives the correlation observed in Japanese. The remainder of the paper, in section 3, presents new data corroborating predictions of this account for intervention in Japanese.

2 Analysis

2.1 Kotek's (2017) proposal in a nutshell

Kotek (2017) proposes that intervention effects are due to a logical problem (described below) that occurs when any quantifier takes scope between a *wh*-phrase and C at LF:[4]

(9) **Intervention is the result of scope-taking across focus (Kotek 2017):**
Movement into a scope position above *wh*-in-situ at LF leads to ungrammaticality.

(10) **Kotek's intervention schema:**

* LF: C ... **λ** ... *wh*

[3] We follow Shimoyama (2011) in analyzing *wh-mo* NPIs as wide-scope ∀ over negation.

[4] Throughout, arrows indicate movement, and squiggly arrows indicate areas of in-situ (alternative) computation. These arrows are used as a notational convenience only.

That is, whether or not a quantifier acts as an intervener depends on whether or not it can *move out of the way* at LF to avoid the configuration in (10). We assume that *wh*-phrases can be interpreted in-situ at LF by introducing Rooth-Hamblin alternatives which compose pointwise (squiggly arrow) and which will be interpreted by the interrogative complementizer; see e.g. Beck (2006) and Kotek (2017) for details.

Previous literature on focus and *wh* semantics has recognized a problem with defining Predicate Abstraction (PA) over sets of alternatives in simple semantic models (Rooth 1985, Shan 2004, Novel and Romero 2009, Ciardelli et al. 2017; see also Poesio 1996, among others). In brief, standard syncategorematic PA rules (as in Heim and Kratzer 1998) are not well-defined over sets of alternatives. PA over a set of propositional alternatives should intuitively apply *pointwise*, yielding *a set of functions*. However, because the input to PA is an assignment-sensitive set of propositions, PA yields instead *a function returning a set of propositions*.

Shan (2004) demonstrates that simple solutions assumed in the previous literature — transposing a *function into sets of propositions* that a PA rule yields into a *set of functions*, using a type-shifter as in (11) — leads to a problem of over-generation. The result includes both (desired) constant functions (12) but also (undesired) non-constant ones (13).

(11) **A type-shifter for turning type $\langle e, \langle \tau, t \rangle \rangle$ functions into type $\langle \langle e, \tau \rangle, t \rangle$ sets:**
$\lambda Q_{\langle e, \langle \tau, t \rangle \rangle} \ . \ \{ \ f_{\langle e, \tau \rangle} : \forall x_e \ . \ f(x) \in Q(x) \ \}$

(12) **Constant $\langle e, t \rangle$-functions**
$$\left\{ \begin{bmatrix} x_1 \mapsto \text{Alice saw } x_1 \\ x_2 \mapsto \text{Alice saw } x_2 \\ x_3 \mapsto \text{Alice saw } x_3 \end{bmatrix}, \begin{bmatrix} x_1 \mapsto \text{Barbara saw } x_1 \\ x_2 \mapsto \text{Barbara saw } x_2 \\ x_3 \mapsto \text{Barbara saw } x_3 \end{bmatrix}, \begin{bmatrix} x_1 \mapsto \text{Carol saw } x_1 \\ x_2 \mapsto \text{Carol saw } x_2 \\ x_3 \mapsto \text{Carol saw } x_3 \end{bmatrix} \right\}$$

(13) **Non-constant $\langle e, t \rangle$-functions**
$$\left\{ \begin{bmatrix} x_1 \mapsto \text{Alice saw } x_1 \\ x_2 \mapsto \text{Carol saw } x_2 \\ x_3 \mapsto \text{Barbara saw } x_3 \end{bmatrix}, \begin{bmatrix} x_1 \mapsto \text{Alice saw } x_1 \\ x_2 \mapsto \text{Barbara saw } x_2 \\ x_3 \mapsto \text{Carol saw } x_3 \end{bmatrix}, \begin{bmatrix} x_1 \mapsto \text{Carol saw } x_1 \\ x_2 \mapsto \text{Barbara saw } x_2 \\ x_3 \mapsto \text{Alice saw } x_3 \end{bmatrix} \right\}$$

Previous work has proposed instead to type-lift all denotations, either to take assignment functions as arguments (Novel and Romero 2009; see also Poesio 1996), or to operate over sets of propositions (Ciardelli et al. 2017, Charlow 2017), so PA can be defined. Another suggestion is to eschew movement/PA altogether (Shan 2004). In contrast, Kotek argues that this fundamental inability of defining PA over non-trivial sets of alternatives should not be "solved" — instead, it is precisely what gives rise to intervention, (10). We refer the reader to the above-cited works for more details and for additional data.

2.2 Explaining the correlation

Based on the consideration of scope interactions between different quantificational objects and negation in Japanese, Shibata (2015a,b) argues that all objects in Japanese (DP arguments in vP) move overtly out of vP. Objects also necessarily move out of NegP, if present, which Shibata argues has a fixed position just above vP. We further assume the vP-internal subject hypothesis (see e.g. Fukui 1986, Kitagawa 1986, Kuroda 1988), concluding that all (DP) arguments evacuate vP in Japanese. These assumptions are illustrated schematically in (14a). Quantifiers then vary with respect to their ability to reconstruct: those which cannot reconstruct have obligatory wide-scope with respect to negation (14b), whereas those which can reconstruct lead to scope ambiguities with respect to negation, allowing the LFs in (14b) or (14c).

(14) **Scope-taking in Japanese (Shibata 2015a,b):**

a. All arguments move out of vP:
$[_{CP}$ ... DP ... $[_{vP}$ ... t ... V]]

b. LF interpretation in surface position leads to wide scope over negation:
LF: $[_{CP}$ ... DP λx ... $[_{NegP}$ $[_{vP}$ ... x ... V] Neg]] DP > Neg

c. Some (not all) quantifiers reconstruct into vP, allowing narrow scope:
LF: $[_{CP}$ ... $[_{NegP}$ $[_{vP}$ ... DP ... V] Neg]] Neg > DP

Now consider a surface structure where the DP could lead to an intervention configuration (15a). (Movement of the *wh*-phrase to its surface position is not illustrated. The interpreting complementizer is at the left edge of CP for illustration purposes only.) If the quantifier is scope-rigid, it has no choice but to lead to the LF configuration as in (15b). This is a Kotek intervention configuration (10): the calculation of Rooth-Hamblin alternatives must cross an instance of Predicate Abstraction (λx, in bold), which cannot be defined. But if a quantifier is not scope-rigid — i.e. it can reconstruct at LF — the LF in (15c) will also be available. Alternatively, scrambling the *wh*-word above the potential intervener also avoids intervention (15d) without requiring the DP to reconstruct. Finally, the possibility of scoping the quantifier out of the question itself (15e) offers one additional means for avoiding intervention.[5]

(15) **Deriving the generalization (8):**

a. Potential intervener (DP) above *wh*:
$[_{CP}$ C ... DP ... *wh* ... $[_{vP}$ ... t ... V]]

b. *LF interpretation in surface position lead to intervention!*
* LF: $[_{CP}$ C ... DP $\boldsymbol{\lambda x}$ ... *wh* ... $[_{vP}$ ... $\boldsymbol{x}$ ... V]]

c. Reconstruction avoids the intervention configuration:
✓ LF: $[_{CP}$ C ... *wh* ... $[_{vP}$ ... DP ... V]]

d. Scrambling *wh* above also avoids intervention:
✓ LF: $[_{CP}$ C ... *wh* λy ... DP $\boldsymbol{\lambda x}$ ... y ... $[_{vP}$ ... $\boldsymbol{x}$ ... V]]

e. Scoping the quantifier out of the question also avoids intervention:
✓ LF: ... DP λx ... $[_{CP}$ C ... ____ ... *wh* ... $[_{vP}$ ... x ... V]]

3 Predictions of the account

In the remainder of this paper we present three predictions of our account and show that they are indeed borne out by the data, supporting the approach to intervener-hood and intervention presented here. We believe that these findings are not predicted by existing accounts of intervention effects in Japanese.

[5] Note that in order to predict no intervention in cases of reconstruction (15c) and of further movement (15e), all intermediate landing sites of movement — between DP's base position and its final *scope* position at LF — must be ignored as far as the computation of intervention configurations is concerned. Instead, the λ-binder at the final LF position of the moved DP must directly bind its lower variable. See Kotek (2017) for discussion.

3.1 Non-intervention through reconstruction

First, we concentrate on our characterization of *non-intervening* quantifiers. We claim that quantifiers which descriptively do not intervene can do so by reconstructing into a lower, *v*P-internal base position. Therefore in a potential intervention configuration, we predict that the potentially intervening quantifier must be *interpreted* in this reconstructed position inside *v*P.

We first test this forced reconstruction by considering the scope of the intervening quantifier with respect to sentential negation. Following Futagi (2004), we showed in Erlewine and Kotek 2017 that the *only* particle *dake* inside a postposition (DP-*dake*-P) can take scope above or below sentential negation, and at the same time is descriptively a non-intervener. Now consider example (16) below. The quantificational PP 'with only Hanako' *Hanako-dake-to* is in a higher position than the *wh*-word in the surface structure, so we predict that it will be forced to reconstruct into its *v*P-internal base position, which will necessarily be below negation.

(16) **DP-*dake*-P must reconstruct below *wh*; *only* > *not* reading is not possible:**

Taro-wa Hanako-**dake**-to *nani*-o tabe-**nai**-no?
Taro-TOP Hanako-only-with what-ACC eat-NEG-Q

a. * 'What does Taro only not eat with Hanako$_F$?' only > not
Answer: Squid ink pasta (because he gets embarrassed)

b. ? 'What does Taro not eat with only Hanako$_F$?' not > only
Answer: Dimsum (because it's better with more people)

The two potential readings are illustrated by the potential expected answers and respective contexts: what is x such that, just when he is with Hanako, Taro won't eat x (wide scope for *only* over negation), *vs* what is x such that Taro does not eat x with Hanako alone (narrow scope for *only*). While both readings are plausible in appropriate supporting contexts, and *-dake*-P can generally scope above or below negation, only (16b) is possible here. This is as predicted by the reconstruction account of non-intervention, illustrated in (15c) above.

We note that scrambling the *wh*-word above *Hanako-dake-to* makes both readings available. This, too, is predicted by our account. See the LF schema in (15d).

(17) **When *wh* scrambles above intervener, both scope readings become available:**

Taro-wa *nani*-o Hanako-**dake**-to ___ tabe-**nai**-no?
Taro-TOP what-ACC Hanako-only-with ___ eat-NEG-Q

a. ? 'What does Taro only not eat with Hanako$_F$?' only > not

b. ? 'What does Taro not eat with only Hanako$_F$?' not > only

Next, consider the collective vs distributive event interpretation of subjects. We assume that distributive readings require a short movement of the subject. Example (18) provides a baseline, illustrating that in the absence of an intervener, universally quantified subjects in Japanese allow for both collective and distributive interpretations. However, when these quantifiers c-command an in-situ *wh*-phrase, only a collective interpretation is possible, (19).

(18) **Baseline: collective and distributive readings with *zen'in*:**

[Gakusei **zen'in**]-ga LGB-o ka-tta.
student all-NOM LGB-ACC buy-PAST

a. 'All the students together bought a copy of LGB.' collective

b. 'All the students each bought a copy of LGB.' distributive

(19) ***Zen'in* must reconstruct below *wh*; only the collective reading survives:**

[Gakusei **zen'in**]-ga [*dono hon*]-o ka-tta-no?
student all-NOM which book-ACC buy-PAST-Q

a. ✓ 'Which book(s) did the students all buy together?' collective
b. * 'Which book(s) did the students all individually buy?'
(and they each bought other books too) distributive

Here too, scrambling the *wh*-phrase above the quantifier allows for both the collective and distributive readings (20). The distributive reading is possible in (20) because scrambling the *wh*-phrase higher (15d) makes it no longer necessary to reconstruct the quantifier (15c) in order to interpret the *wh*-question.

(20) **When *wh* is scrambled above zen'in, both readings are again available:**

[*Dono hon*]-o [gakusei **zen'in**]-ga ___ ka-tta-no?
which book-ACC student all-NOM buy-PAST-Q

a. ✓ 'Which book(s) did the students all buy together?' collective
b. ✓ 'Which book(s) did the students all individually buy?' distributive

3.2 Non-intervention by scoping out

Next, we consider another way of avoiding intervention, discussed in prior literature for German in Beck 1996 and for English in Pesetsky 2000 and Kotek 2014: A quantifier can avoid causing an intervention effect if it is able to scope out of the question and quantify-in, see (15e). This is possible with universal quantifiers, and leads to a predicted wide-scope reading of the quantifier with respect to the *wh*-phrase — a pair-list reading (see e.g. Karttunen 1977, Comorovski 1989, 1996, É Kiss 1993, Krifka 2001).

The relevant example in given in (21). The embedded question in (21) allows the collective interpretation but not a distributive interpretation, just as in (19) above. However, this sentence has another reading where *all students* takes wide scope out of the question. The resulting interpretation, then, expects that each student bought a (potentially different) book, and that this *list of pairs* is what the teacher would like to know.[6]

(21) **An additional possible reading: A pair-list with *zen'in* quantifying-in**

Sensei-wa [$_{CP}$ [gakusei **zen'in**]-ga [*dono hon*]-o ka-tta-ka] shiri-tai.
teacher-TOP student all-NOM which book-ACC buy-PAST-Q know-want

'The teacher wants to know...

a. ✓ [which book(s) the students all bought together].' collective (19a)
b. * [which book(s) the students all individually bought].' distributive (19b)
c. ✓ [for each student$_i$, which book(s) they$_i$ bought].' pair-list

[6] Matrix questions with universal quantifiers also permit pair-list interpretations, but this reading seems clearer at least in this example when embedded, as in (21).

3.3 Base-generated quantifiers are not interveners

Finally, we return again to the fact that the proposal above ties intervention to movement into a position between the in-situ *wh* and C. The data we have seen so far is compatible with the interpretation of *wh*-in-situ being interrupted by (a) *any* quantification or (b) λ-binders of quantifiers in *derived* positions. Here we offer an argument to tease these two potential explanations apart.

Our proposal predicts that quantifiers that are base-generated high and can be interpreted in their base positions would not be interveners.[7] In example (22), this is shown to be the case using the adjunct 'only on Tuesdays,' which unlike arguments, can be base-generated in a high position and does not require movement out of a low vP position (see section 2.1).

(22) **Temporal modifiers base-generated high do not cause intervention:**

✓ Taro-wa kayoubi-ni-**dake** *nani*-o tabe-ru-no?
Taro-TOP Tuesday-on-ONLY what-ACC eat-NONPAST-Q
'What does Taro eat only on Tuesdays?'

We observe that this adjunct does not cause an intervention effect, supporting hypothesis (b), that it is specifically quantificational material interpreted in a derived position that triggers intervention, over hypothesis (a), that simply any quantificational material triggers intervention.

4 Conclusion

Intervention effects have been the subject of a large and growing body of literature over the past 30 years. Previous work offered rigid descriptions of the set of interveners — be it as related to the semantics of focus (Kim 2002, Beck 2006, Beck and Kim 2006), quantification (Beck 1996), topichood (Grohmann 2006), prosody (Tomioka 2007), (anti-)additivity (Mayr 2014), or semantic type-mismatch (Li and Law 2016). We argued here that these descriptions will all necessarily fall short of the desired result.

Instead, we argued that intervener-hood is crucially tied to a (potential) intervener's scope position at LF: Following Kotek 2017, interveners are those elements which *move* into a scope position that separates an in-situ *wh*-phrase from the interrogative complementizer that must interpret it at LF, and which *cannot* move out of the way. A (potential) intervener can evade intervention by moving out of the way in one of two ways: (a) some quantifiers are able to reconstruct to a base-position below *wh*-in-situ, and (b) some quantifiers are able to scope above interrogative C and quantify into the question. In addition, as has been widely observed, *wh*-in-situ can evade intervention through scrambling above the intervener. We conclude that all DPs in a derivation act as potential interveners, and their precise nature as interveners or non-interveners in a particular derivation will be tied to their possible syntactic positions at LF and the reflexes of their interpretation. It follows that the goal of a theory of intervention is not to pre-classify quantifiers as interveners or non-interveners, but instead to consider the scope-taking possibilities of all potential interveners.

[7] We thank Paloma Jeretič (p.c.) for suggesting this prediction and to Yohei Oseki (p.c.) for initial discussion.

References

Beck, Sigrid. 1996. Quantified structures as barriers for LF movement. *Natural Language Semantics* 4:1–56.

Beck, Sigrid. 2006. Intervention effects follow from focus interpretation. *Natural Language Semantics* 14:1–56.

Beck, Sigrid, and Shin-Sook Kim. 2006. Intervention effects in alternative questions. *Journal of Comparative German Linguistics* 9:165–208.

Charlow, Simon. 2017. The scope of alternatives: Indefiniteness and islands. Manuscript, Rutgers University.

Ciardelli, Ivano, Floris Roelofsen, and Nadine Theiler. 2017. Composing alternatives. *Linguistics and Philosophy* 40:1–36.

Comorovski, Ileana. 1989. Discourse and the syntax of multiple constituent questions. Doctoral Dissertation, Cornell University.

Comorovski, Ileana. 1996. *Interrogative phrases and the syntax-semantics interface*. Dordrecht: Kluwer.

É Kiss, Katalin. 1993. Wh-movement and specificity. *Natural Language & Linguistic Theory* 11:85–120.

Erlewine, Michael Yoshitaka, and Hadas Kotek. 2017. Intervention tracks scope-rigidity in Japanese. In *Proceedings of LENLS 14*.

Fukui, Naoki. 1986. A theory of category projection and its application. Doctoral Dissertation, Massachusetts Institute of Technology.

Futagi, Yoko. 2004. Japanese focus particles at the syntax-semantics interface. Doctoral Dissertation, Rutgers, The State University of New Jersey.

Grohmann, Kleanthes K. 2006. Top issues in questions: Topics—topicalization—topicalizability. In *Wh-movement: Moving on*, ed. Lisa Lai-Shen Cheng and Norbert Corver. Cambridge, MA: MIT Press.

Hasegawa, Nobuko. 1995. *Wh*-gimonbun, hitei-taikyoku-hyogen-no *shika*, to *also* no *mo* [*wh*-questions, NPI *shika*, and 'also' *mo*]. In *Proceedings of the Third International Nanzan University Symposium on Japanese Language Education and Japanese Linguistics*, 107–128.

Heim, Irene, and Angelika Kratzer. 1998. *Semantics in generative grammar*. Malden, Massachusetts: Blackwell.

Hoji, Hajime. 1985. Logical form constraints and configurational structures in Japanese. Doctoral Dissertation, University of Washington.

Karttunen, Lauri. 1977. Syntax and semantics of questions. *Linguistics and Philosophy* 1:3–44.

Kataoka, Kiyoko. 2006. Neg-sensitive elements, neg-c-command, and scrambling in Japanese. In *Japanese/Korean Linguistics 14*, 221–233.

Kim, Shin-Sook. 2002. Intervention effects are focus effects. In *Japanese/Korean Linguistics 10*, 615–628.

Kitagawa, Yoshihisa. 1986. Subjects in Japanese and English. Doctoral Dissertation, University of Massachusetts Amherst.

Kotek, Hadas. 2014. Composing questions. Doctoral Dissertation, Massachusetts Institute of Technology.

Kotek, Hadas. 2017. Intervention effects arise from scope-taking over alternatives. In *Proceedings of NELS 47*, ed. Andrew Lamont and Katerina Tetzloff, volume 2, 153–166. Amherst, MA: GLSA.

Krifka, Manfred. 2001. Quantifying into question acts. *Natural Language Semantics* 9:1–40.

Kuroda, Sige-Yuki. 1988. Whether we agree or not: a comparative syntax of English and Japanese. *Linguistic Investigations* 12:1–47.

Li, Haoze, and Jess Law. 2016. Alternatives in different dimensions: A case study of focus intervention. *Linguistics and Philosophy* 39:201–245.

Mayr, Clemens. 2014. Intervention effects and additivity. *Journal of Semantics* 31:513–554.

Mogi, Toshinobu. 2000. Toritate-shi-no kaisosei-ni tsuite [on the layeredness of focus particles]. In *Proceedings of the Fall 2000 meeting of the Society for Japanese Linguistics*, 54–61.

Novel, Marc, and Maribel Romero. 2009. Movement, variables, and Hamblin alternatives. In *Proceedings of Sinn und Bedeutung 14*.

Pesetsky, David. 2000. *Phrasal movement and its kin*. Cambridge, MA: MIT Press.

Poesio, Massimo. 1996. Semantic ambiguity and perceived ambiguity. In *Semantic ambiguity and underspecification*, ed. Kees van Deemter and Stanley Peters, chapter 8, 159–201. Chicago, IL.: CSLI Publications.

Rooth, Mats. 1985. Association with focus. Doctoral Dissertation, University of Massachusetts, Amherst.

Shan, Chung-chieh. 2004. Binding alongside Hamblin alternatives calls for variable-free semantics. In *Proceedings of SALT 16*.

Shibata, Yoshiyuki. 2015a. Exploring syntax from the interfaces. Doctoral Dissertation, University of Connecticut.

Shibata, Yoshiyuki. 2015b. Negative structure and object movement in Japanese. *Journal of East Asian Linguistics* 24:217–269.

Shimoyama, Junko. 2011. Japanese indeterminate negative polarity items and their scope. *Journal of Semantics* 28:413–450.

Takahashi, Daiko. 1990. Negative polarity, phrase structure, and the ECP. *English Linguistics* 7:129–146.

Tomioka, Satoshi. 2007. Pragmatics of LF intervention effects: Japanese and Korean interrogatives. *Journal of Pragmatics* 39:1570–1590.

Yanagida, Yuko. 1996. Syntactic QR in *wh-in-situ* languages. *Lingua* 99:21–36.

Cross-linguistic evidence for a non-distributive lexical meaning of conjunction *

Enrico Flor[1], Nina Haslinger[1], Hilda Koopman[2], Eva Rosina[1], Magdalena Roszkowski[1], and Viola Schmitt[1]

[1] University of Vienna, Vienna, Austria
enrico.flor@univie.ac.at
nina.haslinger@univie.ac.at
eva.rosina@univie.ac.at
magdalena.roszkowski@univie.ac.at
viola.schmitt@univie.ac.at
[2] UCLA, Los Angeles, California, U.S.A.
koopman@ucla.edu

Abstract

This paper investigates the lexical meaning of elements like English *and* ('COORD') in conjunctions with individual-denoting conjuncts by considering cross-linguistic form-function correlations. We present two generalizations concerning the correspondence between distributive readings and formal markedness both inside and outside the coordinate structure. We argue that they suggest that the cross-linguistic lexical meaning of COORD is non-distributive and that distributivity is introduced by additional operators. We then discuss how existing semantic treatments of coordinate structures could be adapted to yield a compositional analysis of the cross-linguistic facts.

1 Introduction

This paper focusses on the lexical meaning of English *and* and its correlates in other languages ('COORD') in '*e*-conjunctions' – conjunctions with individual-denoting conjuncts as in (1-a).

(1) a. *Mary and Sue earned exactly 100 euros.*
b. 'Mary earned exactly 100 euros and Sue earned exactly 100 euros.' D-reading
c. 'Mary and Sue earned exactly 100 euros between them.' ND-reading

What we will call **D-theories** assume that this lexical meaning is **distributive**, ('D'), which roughly means that it is reducible to the operation '$\wedge$' from classical propositional logic. **ND-theories**, on the other hand, take it to be **non-distributive** ('ND') which essentially means that COORD expresses an operation analogous to that which forms pluralities from individuals. Each type of analysis has to assume additional operations to derive certain readings of sentences with *e*-conjunctions: For (1-a), D-theories require additional operations to derive the ND-reading in (1-c), whereas ND-theories need additional operations for the D-reading in (1-b).

*We thank Moreno Mitrović and Uli Sauerland for comments and discussion. We would also like to thank our consultants Nikolaos Angelopoulos, Paul Roger Bassong, Zhuo Chen, Jovana Gajić, Cristina Guardiano, Emily Hanink, Soohwan Jung, Travis Major, Pam Munro, Edgar Onea, Sozen Ozkan, Augustina Owusu, Zixian Qiu, Yasu Sudo and Marcin Wągiel for their contributions to our Terraling group 'Conjunction and Disjunction' (cf. http://test.terraling.com/groups/8). All errors are our own. This research was funded by the Austrian Science Fund (FWF), project P 29240-G23, 'Conjunction and disjunction from a typological perspective'.

But does cross-linguistic evidence support either of these two theories? Based on data from both the literature and our own ongoing study, we address this question by looking at form-function correlations: Broadly speaking, we try to relate the different additional operations posited by the two theories to overt morphological markers that appear in sentences with *e*-conjunctions. Two formal properties will be correlated with the availability of D- and ND-readings: First, we will look at paradigms of **conjunction patterns**, i.e. ways of expressing the coordinate structure itself. In these paradigms the formal realization of COORD is held constant, but the coordinate structure may contain additional conjunction particles (α), as schematized in (2-a). Second, we will examine paradigms of **conjunction strategies**, which also contain material outside the coordinate structure. In these paradigms, the coordinate structure itself is held constant and variation concerns additional markers (β) occurring *outside* the coordinate structure (2-b).[1] Our survey is restricted in three respects: First, we only consider instances of *iterative coordination*, i.e. coordinate structures that allow for more than two conjuncts (cf. [2] for a more precise definition). Second, we only look at *e*-conjunctions occurring in *subject position*. Third, we only investigate sentences where such conjunctions occur with what we call **C-predicates**, namely, predicates containing a degree expression, as in (1-a) above, or an indefinite plural, e.g. *read exactly five books*.[2]

(2) a. [A COORD B] [P] vs. [A COORD B α] [P]
b. [A COORD B] [P] vs. [A COORD B] [β P]

Based on our (limited) data set, we present two generalizations which, if they should turn out to be cross-linguistically valid (among languages that have iterative *e*-conjunction in the first place), would have the following theoretical consequences: First, the lexical meaning of COORD is ND. Second, at least one of the additional operations required by ND-theories – so-called VP-level distributivity operators – must be available cross-linguistically. We then investigate how our findings can be implemented. In most existing theories of COORD, its lexical meaning is defined as a *binary* operation on the conjuncts' denotations. However, many languages display conjunction patterns where *each* conjunct is morphologically marked by α, [6, 10]. This suggests that, in addition to COORD, some conjunction patterns involve a *unary* operator modifying each conjunct. We show how the semantic analyses of such structures in [6, 10] could be adapted to a ND lexical meaning for COORD and point to the remaining problems.

2 Background: Theories of conjunction

D-analyses of conjunction (cf. e.g. [8]) take the meaning of COORD to be defined in a unified way as in (3-b) for all types that "end in *t*", (3-a), thus accounting for the cross-categorial applicability of COORD in languages like English. For *e*-conjunctions we thus need the operation in (3-c) that shifts the denotations of the conjuncts to a *t*-conjoinable type. As a result, we derive the D-reading of *e*-conjunctions like (1-a) as in (3-e), using the derived meaning for quantifier conjunction in (3-d).

(3) a. The set TC of *t*-conjoinable types is the smallest set of semantic types such that

[1] Neither of the schemata in (2) is supposed to represent linear order facts or the number of occurrences of α/β. Nor do we assume that COORD must be phonologically realized.

[2] We explicitly instructed our consultants to use C-predicates, and to avoid sentences with inherently distributive predicates – such as *John, Mary and Sue left* – since such predicates won't let us distinguish the D-reading and the ND-reading truth-conditionally. Therefore, our claims about the presence and the obligatoriness of certain distributivity markers might not generalize to inherently distributive predicates.

$t \in TC$ and if $b \in TC$, then for all a, $\langle ab \rangle \in TC$. (cf. [8])

b. $[\![\text{COORD}_t]\!] = \lambda p_t.\lambda q_t.p \wedge q$, and for every type $b \in TC$ and every type a: $[\![\text{COORD}_{\langle ab \rangle}]\!] = \lambda P_{\langle ab \rangle}.\lambda Q_{\langle ab \rangle}.\lambda x_a.[\![\text{COORD}_b]\!](P(x))(Q(x))$ (cf. [8])

c. $[\![\uparrow]\!] = \lambda x_e.\lambda P_{\langle et \rangle}.P(x)$ (cf. [7])

d. $[\![\text{COORD}_{\langle\langle et \rangle t \rangle}]\!] = \lambda \mathbf{P}_{\langle\langle et \rangle t \rangle}.\lambda \mathbf{Q}_{\langle\langle et \rangle t \rangle}.\lambda R_{\langle et \rangle}.\mathbf{P}(R) \wedge \mathbf{Q}(R)$

e. $[\![[\uparrow \text{M}_e]\ \text{COORD}_{\langle\langle et \rangle t \rangle}\ [\uparrow \text{S}_e]]\ \text{P}]\!] = [\![\text{P}]\!]([\![\text{M}]\!]) \wedge [\![\text{P}]\!]([\![\text{S}]\!])$

Without further assumptions, the D-analysis does not straightforwardly account for the ND-reading in (1-c). Yet, D-analyses *can* retrieve ND-readings by assuming additional operations. [11] posits two operators, MIN (4-a), and $\exists$ (4-b), which attach to the conjunction (we slightly adapt his proposal for our purposes). In combination, they yield existential quantification over those pluralities[3] consisting exclusively of individuals the conjuncts' denotations identify (4-d) – which will give us the ND-reading for sentences like (1-a).

(4) a. $[\![\text{MIN}]\!] = \lambda \mathcal{P}_{\langle\langle et \rangle t \rangle}.\lambda x_e.\exists Q_{\langle et \rangle}.[\mathcal{P}(Q) \wedge \forall Q'_{\langle et \rangle}[Q' \subseteq Q \wedge \mathcal{P}(Q') \rightarrow Q' = Q] \wedge x = \bigoplus Q]$

b. $[\![\exists]\!] = \lambda P_{\langle et \rangle}.\lambda Q_{\langle et \rangle}.\exists x_e[P(x) \wedge Q(x)]$

c. [[$\exists$ [MIN [[$\uparrow$ Mary] $\text{COORD}_{\langle\langle et \rangle t \rangle}$ [$\uparrow$ Sue]]]] [earned 100 euros]]

d. $[\![\exists\ [\text{MIN}\ [[\uparrow \text{M}]\ \text{COORD}_{\langle\langle et \rangle t \rangle}[\uparrow \text{S}]]]]\!] = \lambda Q_{\langle et \rangle}.\exists x_e[x = m \oplus s \wedge Q(x)] = \lambda Q_{\langle et \rangle}.Q(m \oplus s)$

ND-analyses of conjunction (cf. e.g. [4]) on the other hand assume that COORD denotes a sum operation ('$\oplus$') in the individual domain, (5-a). e-conjunctions denote pluralities of individuals and we straightforwardly derive the ND-reading. With this type of analysis, additional operations – e.g. a **distributivity operator** – are required to derive the D-reading. There are two potential implementations: D_1 in (5-b) shifts a type e plurality to a distributive quantifier. Applying D_1 after COORD yields the same result as the D-analysis in (3-e). D_2 in (5-c) modifies the predicate rather than the subject (cf. a.o. [5]). (6) shows that both approaches yield the same result for (1-a), but as shown below, they make distinct cross-linguistic predictions.

(5) a. $[\![\text{COORD}_e]\!] = \lambda x_e.\lambda y_e.x \oplus y$

b. $[\![\text{D}_1]\!] = \lambda x_e.\lambda P_{\langle et \rangle}.\forall y \leq_{AT} x.P(y) = 1$

c. $[\![\text{D}_2]\!] = \lambda P_{\langle et \rangle}.\lambda x_e.\forall y \leq_{AT} x.P(y) = 1$

(6) a. $[\![[\text{D}_1$ [Mary COORD_e Sue]] earned 100 euros$]\!] = [\lambda P_{\langle et \rangle}.\forall y \leq_{AT} m \oplus s.P(y) = 1]$ $([\![\text{earned 100 euros}]\!]) = 1$ iff $\forall y \leq_{AT} m \oplus s.[\![\text{earned 100 euros}]\!](y) = 1$

b. $[\![[[\text{Mary COORD Sue}]\ [\text{D}_2\ [\text{earned 100 euros}]]]]\!] = [\lambda x_e.\forall y \leq_{AT} x.[\![\text{earned 100 euros}]\!](y) = 1](m \oplus s) = 1$ iff $\forall y \leq_{AT} m \oplus s.[\![\text{earned 100 euros}]\!](y) = 1$

Our question in the following will be whether one of the analyses could hold **universally** (among languages with iterative e-conjunctions).[4]

[3] We assume a set $A \subseteq D_e$ of atomic individuals, a binary operation $\oplus$ on D_e and a function $f : (\mathcal{P}(A) \setminus \{\emptyset\}) \rightarrow D_e$ such that: 1) $f(\{a\}) = a$ for any $a \in A$ and 2) f is an isomorphism between the structures $(\mathcal{P}(A) \setminus \{\emptyset\}, \cup)$ and $(D_e, \oplus)$. Hence there is a one-to-one correspondence between plural individuals and nonempty sets of atomic individuals. We will use the notions in (i), following much of the literature.

(i) For any $a, b \in D_e$, $S \subseteq D_e$:

a. $a \leq b \Leftrightarrow a \oplus b = b$ ("a is a part of b")

b. $a \leq_{AT} b \Leftrightarrow a \leq b \wedge a \in A$ ("a is an atomic part of b")

c. $\bigoplus S = f(\bigcup\{f^{-1}(x) \mid x \in S\})$ (the sum of all individuals in S)

[4] We rule out the possibility that COORD is lexically ambiguous between a D-meaning and a ND-meaning: This is unlikely to be universally correct, given examples like (i) (adapted from [1]) which show that at least

3 Correlating form and function cross-linguistically

As the two analyses differ in which reading of sentences like (1) they take to be 'basic', they make different predictions about cross-linguistic form-function correlations. These relate to how formal *markedness* relations between coordination patterns or strategies correlate with distributivity. We present two cross-linguistic generalizations, one about coordination *patterns*, one about coordination *strategies*. We then specify the predictions of the analyses and show that the generalizations support the ND-analysis for a cross-linguistic lexical meaning of COORD.

Our data set comprises examples from the literature and from our on-going Terraling study 'Conjunction and disjunction' which currently contains data from 15 languages.[5] Terraling is an open-ended, open-source database where language experts (mostly native speaker linguists) answer metalinguistic questions in a 'yes/no/does-not-apply' format, and also have the option of providing glossed examples (cf. [3]). Our study is the first to use this database for formal semantics. Therefore, in our questionnaire we asked consultants whether particular forms were available in their language (with a focus on the presence / absence of additional markers that enforce a certain reading), but importantly we also asked whether these different forms can express D- and ND-readings [2]. In particular, we asked our consultants to construct sentences with C-predicates for different coordination patterns/strategies in their language. They then had to test for the presence of distributive and non-distributive interpretations by judging whether these sentences adequately describe certain scenarios that distinguish between the two readings. We asked consultants to use modified numerals inside the C-predicate where possible, in order to make it easier to distinguish the two readings truth-conditionally.

3.1 Generalization A: Conjunction patterns

Generalization A concerns markedness relations within the coordinate structure itself.

(**A**) **Generalization A**: For any pair of iterative coordination patterns within a language that have a conjunctive meaning and apply to proper names, where one pattern can be obtained from the other by adding "additional markers":
(a) If the **marked** pattern permits a **ND** interpretation, so does the **unmarked** pattern.
(b) If the **unmarked** pattern allows for a **D** interpretation, so does the **marked** pattern.

For two *coordination patterns* P and P+α, where P has both a D-reading and a ND-reading and α stands for one or more overt morphological markers inside the coordinate structure, there are three logical possibilities.[6] The first possibility is that P+α could also have both readings, in

in some cases, the ambiguity is due to the predicate rather than COORD: (i) is ambiguous between a D- and a ND-reading of VP2 and can thus be true in a scenario where Mary and Sue drank exactly one glass each. For this reading, a distributive lexical meaning of COORD would be needed – but this conflicts with the requirement that COORD must be non-distributive to license the collective predicate in VP1.

(i) *Mary and Sue [$_{VP1}$ met in the bar] and [$_{VP2}$ had exactly one glass of wine].*

[5] So far, we have data from Akan (Niger-Congo, Kwa), Basa'a (Niger-Congo, Bantu), Cantonese (Sino-Tibetan, Chinese), Chickasaw (Muskogean), Dutch (Indo-European, Germanic), German (Indo-European, Germanic), Greek (Indo-European, Greek), Italian (Indo-European, Italic), Japanese (Japonic), Korean (Koreanic), Nones (Indo-European, Italic), Polish (Indo-European, Balto-Slavic), Serbo-Croatian (Indo-European, Balto-Slavic), Turkish (Turkic) and Wuhu Chinese (Sino-Tibetan, Chinese).

[6] A language may also have two coordination patterns that are not related in an obvious way, i.e. neither of the patterns formally 'contains' the other, e.g. German *A und B* vs. *sowohl A als auch B* 'A as well as B'. Taken at face value, such cases are uninformative w.r.t. our initial question, but cf. [2] for discussion.

which case the additional material α would not affect (non-)distributivity. This case, discussed by [10] for Japanese *A-to B* and *A-to B-to*, is uninformative for the question at hand.

(7) *A-to B(-to) de 100 kg ni naru.*
'A and B weigh 100 kg.' (Japanese ([10, 182, (48)]), both D-and ND-reading available)

The second option is that P+α has only a D-reading, i.e. the additional marking α 'removes' the ND-interpretation. This is exemplified by (8) from Serbo-Croatian (cf. also [10] for Hungarian). The marked pattern P+α in (8-b) 'contains' the unmarked pattern P in (8-a): Whereas in (8-b) the marker *i* modifies each conjunct, this is not the case in (8-a). P is ambiguous between a D-reading and a ND-reading because (8-a) is true in both scenarios in (9). P+α, on the other hand, has only a D-interpretation, because (8-b) is not true in the scenario in (9-b).

(8) a. *[A (i) B i C] su zaradili tačno sto evra.*
A (and) B and C AUX.3PL earn.PART.PL.M exactly hundred euros.GEN
'A, B and C earned exactly 100 euros.'
b. *[I A i B i C] su zaradili tačno sto evra.*
and A and B and C AUX.3PL earn.PART.PL.M exactly hundred euros.GEN
'A, B and C earned exactly 100 euros each.'
(Serbo-Croatian, adapted from examples by Jovana Gajić[7])

(9) a. A earned 100 euros, B earned 100 euros, C earned 100 euros.
b. A earned 30 euros, B earned 30 euros, C earned 40 euros.

The third possibility is that P+α has only a ND-reading. This possibility – excluded by (**A**) – is not attested in our data set, although our survey explicitly asks for examples of this kind. We conjecture that additional marking inside the coordinate structure never 'removes' a D-reading.

(**A**) captures another interesting gap in our data set: It is never the case that P only has a D-reading and P+α has both a D-reading and a ND-reading. It seems that marking inside the coordinate structure never 'adds' a ND-reading. While we did not explicitly ask our consultants whether this pattern exists, we did ask them to provide examples of *e*-conjunctions that only have a D-reading, and of *e*-conjunctions that are ambiguous. These examples never show the markedness relation just described.

3.2 Generalization B: Conjunction strategies

Generalization B relates to additional marking *outside* of the coordinate structure, i.e. on the predicate.

(**B**) **Generalization B**: There are iterative conjunction patterns that require one or more predicate-level markers for a **distributive** interpretation of C-predicates. This means that the D-reading of sentences with a C-predicate is available with the markers, but unavailable if the markers are omitted.
There are no iterative conjunction patterns that require predicate-level markers for a **non-distributive** interpretation of C-predicates.

We are now comparing *coordination strategies* S and S+β, where β stands for one or more additional marker(s) *outside* the coordinate structure and the coordinate structure itself is the same in both strategies. The picture here is analogous to that of coordination patterns: While

[7] *http://test.terraling.com/groups/8/examples/16182, http://test.terraling.com/groups/8/examples/16177*

many languages have overt predicate-level markers that force a ND-interpretation (e.g. English *together*), our data set involves no cases where a predicate-level marker is *required* for a ND-interpretation of a C-predicate. Yet, we do find languages where additional marking on the predicate is required for a D-interpretation, i.e. where S has only a ND-interpretation and S+β allows for a D-interpretation. This is exemplified by Basa'a in (10).

(10) a. *⌈A, B ni C⌋ ɓá-bí-kosná dikóó dísámal*
A B COORD C 2.SM-PST2-receive 13.thousands 13.six
'A, B and C received six thousand francs.' (ND only)

b. *⌈A, B ni C⌋ ɓá-bí-kosná dikóó dísámal, híkií mut*
A B COORD C 2.SM-PST2-receive 13.thousand 13.six each 1.person
'A, B and C received six thousand francs each.' (D only)

(Basa'a, adapted from examples by Paul Roger Bassong[8])

The coordinate structure in (10-a) and (10-b) is the same, but they exemplify different coordination strategies, as *híkií mut* 'each person' is present only in (10-b). The strategy S in (10-a) has only a ND-interpretation and S+β in (10-b) has a D-interpretation.

3.3 Theoretical consequences

While (**A**) and (**B**) are analogous in that some kind of formal 'markedness' is associated with D-interpretations, but not with ND-interpretations, they differ in their theoretical consequences.

As opposed to (**A**), (**B**) relates to formal correlates of the two readings of C-predicates, rather than formal correlates of the two readings of conjunction, since predicate-level D-markers are not part of the coordinate structure itself. Its impact on our initial question concerning the cross-linguistic semantics of COORD is thus indirect – it will help us determine the theoretical consequences of (**A**). Namely, (**B**) suggests that cross-linguistically, the D-interpretation of C-predicates always involves an additional syntactic operator, which is absent in the case of a ND-interpretation. More precisely, we submit that predicate-level operators like D_2 in (5-c) are available in all languages that allow for D-interpretations of C-predicates. Languages differ in whether they have to spell out D_2 overtly: In a language like English in which D_2 is only optionally realized, (11-a) must have a structure like (11-b), with an overt realization of D_2, while (1-a) can correspond to either of the two structures in (11-b) and (11-c). In languages like Basa'a, on the other hand, distributivity operators like D_2 must be realized overtly whenever they are present, and are spelled out as the additional marking β that is needed for a distributive interpretation. In such languages, a sentence with a C-predicate lacking an overt D_2 will thus be unambiguously ND (assuming a ND-interpretation of the coordinate structure itself).

(11) a. *Mary and Sue each earned 100 euros.*
b. [[Mary COORD Sue] [D_2 [earned 100 euros]]]
c. [[Mary COORD Sue] [earned 100 euros]]

With the assumption that predicate level D_2 is indeed available cross-linguistically, we can specify the theoretical predictions of (**A**).[9] Both analyses allow us to derive a ND-meaning for conjunction patterns like English *A, B and C* or Serbo-Croatian *A, B i C* which, when combined with D_2, yields a D/ND ambiguity. To derive conjunction patterns that lack the ND-reading, we have to add an operator like D_1 (12-a). In a language like English, there is

[8] *http://test.terraling.com/groups/8/examples/16284*; *http://test.terraling.com/groups/8/examples/16285*

[9] [2] spell out the parameter settings that have an effect on these predictions and lay out the morpho-syntactic assumptions required to derive them without predicting a transparent markedness relation in every language.

no morphosyntactic evidence for this operator. But in languages where this operator always has an overt morphological reflex, but does not affect the morphological spell-out of COORD, we would get an additional marker that removes the ND-reading – a situation that seems to be attested in several languages including Serbo-Croatian, Hungarian and Turkish (cf. [10] and data on *test.terraling.com/groups/8*) (12).

(12) a. [A COORD B] D or ND (depending on D_2)
b. [D_1 [A COORD B]] D only **ND-analysis**

The D-analysis, on the other hand, predicts that the ND-reading requires the presence of the additional operators MIN and ∃. In this case, if a language required an overt realization of either of these operators, we would get an additional marker that adds the ND-reading to a structure lacking it – a situation unattested in our sample and ruled out by (**A**).

(13) a. [A COORD B] D only
b. [∃ [MIN [A COORD B]]] D or ND (depending on D_2) **D-analysis**

Of course, the predictions of the D-analysis would change if our claim that D_2 is present cross-linguistically turned out to be false. But then we would still predict that additional marking can *remove* the D-interpretation – another unattested situation ruled out by (**A**). In summary, if the unattested markedness relations in our sample reflect real typological gaps, these gaps can be derived from the ND-analysis under certain morphosyntactic assumptions. Further, under the D-analysis the existence of D-only conjunction patterns that are marked relative to an ambiguous conjunction pattern – a situation found in several languages – would be unexpected.

4 Issues for the analysis of conjunction particles

The generalizations above suggest that the lexical meaning of COORD is ND and that a D-reading of the coordination is sometimes derived by means of additional morphology α inside the coordinate structure. So far, the only potential meaning for α we provided was the unary operator D_1 that modified the entire conjunction, but this assumption is at odds with the actual form of the marked patterns that display a D-reading: In several languages a particle – called μ in [6] – is affixed to each conjunct, as witnessed by e.g. (8) above and schematized in (14). This means that μ itself cannot spell out D_1. So how can we compositionally derive a D-reading for (14) while simultaneously maintaining a ND-analysis of COORD?

(14) A-μ COORD B-μ

While [6] and [10] each provide compositional analyses of the formal pattern in (14), neither takes the lexical meaning of COORD to be ND or tries to derive the D-reading from the ND-reading. Hence, neither proposal is compatible with our empirical results and their consequences. In the following, we discuss if these accounts can be adapted to fit our claims above.

4.1 Conjunction particles introduce postsuppositions

Szabolcsi [10] assumes the underlying structure in (15-a) for (14) (adapted here to our examples). The conjuncts must be shifted to a t-conjoinable type (by ↑) and are each affixed by μ, which introduces a postsupposition requiring that the conjunct's denotation is asymmetrically entailed by the denotation of the entire conjunction X. The resulting expressions are then conjoined by COORD, which forms a pair of their denotations (15-b). Finally, the silent operator

$OP_\cap$ applies at the top-level of the conjunction and intersects the elements of the pair, (15-c).

(15) a. $[_X\ OP_\cap\ [_Y\ [[\uparrow \text{Mary}]\ \mu]\ [\textsc{coord}\ [[\uparrow \text{Sue}]\ \mu]]]]$
b. $[\![Y]\!] = \langle [\![\uparrow \text{M}]\!], [\![\uparrow \text{S}]\!] \rangle = \langle \lambda P.P([\![\text{M}]\!]), \lambda P.P([\![\text{S}]\!]) \rangle$ (ignoring postsuppositions)
c. $[\![X]\!] = [\![\uparrow \text{M}]\!] \cap [\![\uparrow \text{S}]\!] = \lambda P_{\langle et \rangle}.P([\![\text{M}]\!]) \wedge P([\![\text{S}]\!])$ (ignoring postsuppositions)

How would we have to modify such a proposal to make it fit our generalizations and their consequences? Recall that we are trying to derive the D-reading of the more complex structure from the ND-reading of the simpler structure in (16). (16) cannot contain $OP_\cap$ (because it has a ND-reading)[10], which in turn means that we have to say something additional about the denotation of (16) to explain how it combines with predicates.

(16) [A [coord B]]

As the meaning of coord should remain constant across the less marked and the more marked patterns, and the lexical meaning of coord in (16) must be ND, we have to generalize the ND-analysis to types ending in t if we want to maintain (15-a) above – otherwise, $OP_\cap$ cannot apply. Furthermore, this generalized ND-analysis must be such that the denotations of the individual conjuncts remain transparent for $OP_\cap$. For this purpose we employ a proposal for generalized sum-formation that is motivated independently in [9]: For any semantic domain D_a there is a set $AT_a \subseteq D_a$ of atomic elements of that domain, a binary operation $\oplus$ on D_a and a function $f_a : (\mathcal{P}(AT_a) \setminus \{\emptyset\}) \to D_a$ such that: 1) $f_a(\{X\}) = X$ for any $X \in AT_a$ and 2) f_a is an isomorphism between the structures $(\mathcal{P}(AT_a) \setminus \{\emptyset\}, \cup)$ and $(D_a, \oplus)$. Assuming that coord occurring with conjuncts of type a always expresses the operation $\oplus$ on D_a, the constituent Y from (15-a) above thus has the denotation in (17-a) – which has the atomic parts $[\![\uparrow \text{M}]\!]$ and $[\![\uparrow \text{S}]\!]$. Accordingly, we have to generalize the denotation of $OP_\cap$ so as to apply to pluralities with arbitrarily many atomic parts, (17-b). Hence, we also derive the meaning in (15-c) for (15-a) above, but our assumptions about the semantic contributions of the individual operators differ from those made by [10].

(17) a. $[\![Y]\!] = [\![\uparrow \text{M}]\!] \oplus [\![\uparrow \text{S}]\!]$
b. $[\![OP_\cap]\!] = \lambda \mathcal{P}_{\langle \langle et \rangle t \rangle}. \bigcap \{\mathcal{Q} : \mathcal{Q} \leq_{AT} \mathcal{P}\}$

In addition, we must posit a syntactic agreement mechanism that ties the occurrence of $OP_\cap$ to that of μ, because we must exclude silent $OP_\cap$ from occurring in structures lacking μ, like (16). If it could apply in these cases, we would falsely predict that languages like Basa'a in (10), where C-predicates are unambiguous, should always allow for D-readings of conjunctions, irrespective of whether the predicate contains a D-marker or not.

Clearly, this adaptation of the proposal is not yet satisfactory. Without additional assumptions concerning the composition of quantifier pluralities with the predicate, $OP_\cap$ seems obligatory whenever coord conjoins expressions of quantifier type and therefore, such conjunctions should be limited to D-readings – but it is well-known that they are not: One of the readings of (18-a) is the ND-reading in (18-b).

(18) a. *Two girls and five boys earned exactly 100 euros.*
b. 'A plurality consisting of two girls and five boys earned exactly 100 euros in total.'

Furthermore, the proposal relies on the availability of $OP_\cap$, which we would expect at least some languages to spell out overtly but which we have not encountered, yet, in our data set.

[10] Adding an additional operator on top of (15-a) which yields the ND-reading is incompatible with (**A**).

4.2 Conjunction particles introduce type-shifts

Mitrović and Sauerland [6] do not posit a silent operator at the top node of the conjunction but rather put the semantic workload on the particles μ and silent morphemes $\Uparrow_{\langle e\langle et\rangle\rangle}$ which each conjunct combines with first (19-a). $\Uparrow_{\langle e\langle et\rangle\rangle}$ maps any individual x to the singleton $\{x\}$, whereas μ shifts expressions from $\langle et\rangle$ to $\langle\langle et\rangle t\rangle$ (19-b). For the meaning of COORD, [6] assume the D-analysis, so that X in (19-a) has the same denotation as (15-c) above.

(19) a. $[_X$ [[Mary $\Uparrow_{\langle e\langle et\rangle\rangle}$] μ] [COORD [[Sue $\Uparrow_{\langle e\langle et\rangle\rangle}$] μ]]]
b. $[\![\mu]\!] = \lambda P_{\langle et\rangle}.\lambda Q_{\langle et\rangle}.P \subseteq Q$

If we want to preserve the structure in (19-a) and the meaning for μ assumed by [6], we have to depart significantly from their analysis of COORD: We require a generalized meaning for COORD that gives us the ND-reading for (16) and also combines with quantifiers. For this purpose, we use a simplified version of the proposal by [9]: Working with the ontology and the denotation for COORD introduced in section 4.1, we add a compositional rule '•' of pointwise application, which applies in two kinds of situations: (i) If an argument plurality $a \oplus b$ combines with a (non-plural) function f that itself does not take pluralities as its argument, the result will be the plurality of values $f(a) \oplus f(b)$. (ii) If a function plurality $f \oplus g$ combines with a (non-plural) argument, the result is again a plurality of values $f(a) \oplus g(a)$. Assuming that sentential pluralities are true iff all of their atomic parts are true, this analysis, partially spelled out in (20-a), correctly derives the D-reading for sentences with the pattern in (19-a). The unmarked pattern in (16), on the other hand, will denote a plurality of individuals ($m \oplus s$). In order to prevent • from applying in (20-b), where the unmarked pattern occurs with a C-predicate that does not contain D_2, we must assume that such predicates primitively take pluralities as their argument and thus combine with the subject plurality by means of functional application. Accordingly, the difference between the marked and the unmarked pattern lies in the type of the coordinates and the assumption that some predicates primitively hold of pluralities.[11]

(20) a. $[\![$[[[[Mary $\Uparrow_{\langle e\langle et\rangle\rangle}$] μ] [COORD [[Sue $\Uparrow_{\langle e\langle et\rangle\rangle}$] μ]]] [earned 100 euros]]$]\!] =$
$(\lambda P_{\langle et\rangle}.\{m\} \subseteq P \oplus \lambda P_{\langle et\rangle}.\{s\} \subseteq P) \bullet [\![$earned 1000 euros$]\!] =$
$= (\{m\} \subseteq [\![$earned 1000 euros$]\!]) \oplus (\{s\} \subseteq [\![$earned 1000 euros$]\!])$
b. $[\![$[[Mary COORD Sue] [earned 100 euros]]$]\!] = [\![$earned 100 euros$]\!]\ (m \oplus s)$

This adaptation also runs into a number of problems. One obvious obstacle is that as in section 4.1, we falsely predict only D-readings for conjunctions in which the conjuncts are of type $\langle\langle et\rangle t\rangle$. Furthermore, our current proposal breaks down in configurations where – according to the assumptions made here – both of the expressions that need to combine with one another denote pluralities, as e.g. in (21) (see [9] for independent arguments supporting this).

(21) *Mary and Sue sang and danced.*

Since (21) has a ND-reading (where it is true if Mary sang and Sue danced), we cannot expand pointwise application to (21). We could introduce a composition rule that combines two pluralities and gives rise to a cumulative reading, as in [9], but given our assumption that COORD forms pluralities cross-categorially, it would be unclear why this mechanism could not apply to quantifier conjunctions like (19-a) and generate a ND-reading for such examples.

[11] Note that as in section 4.1 – and for the very same reasons – we have to make sure that the mechanism associated with the presence of μ is limited to those contexts where it actually occurs. This raises interesting questions wrt. the formal marking of type-shifts in other contexts, which we omit here for reasons of space.

5 Conclusion and Outlook

We considered two theories concerning the lexical meaning of COORD in *e*-conjunctions – the D-analysis and the ND-analysis. Crucially, each of these analyses has to assume additional operations in order to derive some of the readings of sentences containing such conjunctions. We considered whether cross-linguistic formal markedness patterns match the additional operations that each analysis has to posit. We presented two generalizations in our preliminary data set which, in combination, strongly support the ND-analysis. In addition, they suggest that distributivity is the result of additional operations. We then raised the question how we could implement this claim compositionally for structures with conjunction particles. The most interesting empirical questions for future research concern various aspects of the scope of our generalizations: Do (**A**) and (**B**) remain valid once ... (i) we consider more languages? (ii) we expand our data set, e.g. by including *e*-conjunctions in object position? (iii) we consider conjunctions with conjuncts other than individual denoting ones, e.g. predicate conjunctions?

References

[1] David Dowty. A note on collective predicates, distributive predicates and *all*. In *Proceedings of ESCOL 86*, pages 97–115, 1987.

[2] Enrico Flor, Nina Haslinger, Eva Rosina, Magdalena Roszkowski, and Viola Schmitt. Distributive and non-distributive conjunction: Formal semantics meets typology. Ms, University of Vienna, 2017. URL http://www.univie.ac.at/konjunktion/texts.html.

[3] Hilda Koopman. Unifying syntax and morphology and the SSWL database project. Handout for a talk given at The 1st SynCart Workshop, 2016.

[4] Godehard Link. The logical analysis of plurals and mass terms: A lattice-theoretical approach. In R. Bäuerle, C. Schwarze, and A. von Stechow, editors, *Meaning, Use and Interpretation of Language*, pages 302–323. de Gruyter, 1983.

[5] Godehard Link. Generalized quantifiers and plurals. In P. Gärdenfors, editor, *Generalized Quantifiers: Linguistic and Logical Approaches*, pages 151–180. Reidel, Dordrecht, 1987.

[6] Moreno Mitrović and Uli Sauerland. Decomposing coordination. In J. Iyer and L. Kusmer, editors, *Proceedings of NELS 44*, volume 2, pages 39–52, Amherst, 2014. GLSA.

[7] Richard Montague. The proper treatment of quantification in ordinary English. In J. Hintikka, J. Moravczik, and P. Suppes, editors, *Approaches to Natural Language*. Reidel, Dordrecht, 1973.

[8] Barbara Hall Partee and Mats Rooth. Generalized conjunction and type ambiguity. In R. Bäuerle, C. Schwarze, and A. von Stechow, editors, *Meaning, Use and Interpretation of Language*, pages 362–383. de Gruyter, 1983.

[9] Viola Schmitt. Cross-categorial plurality and plural composition. Ms, University of Vienna, 2017. URL http://semanticsarchive.net/Archive/GIwYjFmZ/.

[10] Anna Szabolcsi. What do quantifier particles do? *Linguistics and Philosophy*, 38:159–204, 2015.

[11] Yoad Winter. *Flexibility Principles in Boolean Semantics*. MIT Press, Cambridge, Massachusetts, 2001.

The Gibbard-Harper Collapse Lemma for Counterfactual Decision Theory*

Melissa Fusco[1]

Columbia University, New York City, New York, U.S.A.
mf3095@columbia.edu

Abstract

There is a problem for the debate between causal decision theory, formulated in terms of counterfactuals, and its traditional rival, evidential decision theory: an agent's credences in counterfactuals concerning their own acts collapse into evidential probabilities on those acts once diachronic conditionalization on the act is taken into account. Given assumptions that both classical CDTers [6] and their critics (prominently, [4]) accept, it therefore follows that three things cannot be distinct: (i) the probability of a state, given an act; (ii) the probability that if the act *were* performed, the state *would* result; and (iii) the probability one would have in that same counterfactual, if one learned the act was actually performed.

According to both Evidential and Causal decision theory, a choiceworthy act maximizes expected utility. Evidential decision theory (henceforth 'EDT') recommends that one calculate expected utilities using act-conditionalized probabilities on states. Causal decision theory ('CDT') rejects this strategy, with the rationale that some acts—say, seeking out a therapist—do not cause the states—say, being anxious—with which they are correlated. Allan Gibbard and William Harper's "Counterfactuals and Two Kinds of Expected Utility" [6] provides a classic statement of the causal approach, using counterfactual conditionals to express causal relationships that are uniquely of interest to a CDTer.

Many in the recent literature, however, hold that tides have turned against CDT. One important factor is a 2007 paper by Andy Egan, which presents several counterexamples to the theory.[1] On Egan's view, classical causal decision theorists—in particular, those who appeal to the counterfactual formulation of the theory—adhere to the motto "do whatever has the best expected outcome, holding fixed [one's] initial views about the likely causal structure of the world" (96). However, Egan claims, there are situations where agents should not hold such views fixed. In these cases, agents should anticipate their *future* causal views and leverage those instead, thus taking into account what they will learn by performing the act in question. The force of Egan's point is now widely discussed and widely accepted.[2]

In this paper, I focus on an unappreciated formal result, pointed out by Gibbard & Harper [6] in the third section of their classic paper. There, they prove that if an agent is probabilistically coherent, and the semantics for counterfactuals obeys the principle of the Conditional Excluded Middle [24], then the probability of a counterfactual, given its antecedent, collapses to the probability of the consequent, given the antecedent. It follows that "Future-directed Causal Decision Theory" recommends an act just in case classical *Evidential* Decision Theory does. This complicates the dialectic. By Gibbard & Harper's result, the work of putting causal information into a simple EDT system is exactly *undone* by adding a norm of learning. I present a stronger version of the collapse lemma, which applies to Lewis [19]'s weakening of Gibbard & Harper's axioms. I conclude with some thoughts on how to assess the argument from anticipation in light of these results.

*Warm thanks to Andy Egan, Alex Kocurek, Reuben Stern, two anonymous reviewers for the Amsterdam Colloquium, and the audience at the 2015 Formal Epistemology Workshop for discussion and feedback.

[1] These examples were anticipated by Reed Richter in the 1980s: see Richter [21].

[2] See, for example, Arntzenius [1], Briggs [2] and Greaves [7]. For a prominent dissent, see Joyce [13].

1 EDT vs. CDT (in brief)

Both EDT and CDT begin with the idea that the value of each of a set of available acts (call them the a_i's) is calculated by identifying a set of states which fix one's welfare (a set of states s_k) and then multiplying the utility of each state-act conjunction by one's subjective probability, or *credence*, that that state obtains. This relationship is set out by the expected utility equation in Savage's *Foundations of Statistics*,[3] which applies the generic statistical notion of expectation to the agent's welfare function $Val(\cdot)$. For an arbitrary act a,

$$Val_{\text{Savage}}(a) = \sum_k Cr(s_k)Val(s_k \wedge a) \tag{1}$$

The decision-theoretic maxim then enjoins agents to pick the (or an) act a which maximizes the value of this equation.

A concern about this rule is that states sometimes probabilistically depend on acts. To take a common example, suppose you have a test tomorrow and can either study or party tonight. Partying is more valuable than studying, whether you fail or pass. But you would like to pass, and it is worth more to you to pass than to party. In deliberating, it seems you should take into account that $Cr(\mathsf{pass}|\mathsf{study})$—your subjective probability that you'll pass, *given that you study*—is high, while $Cr(\mathsf{pass}|\mathsf{party})$—your subjective probability that you'll pass, *given that you party*—is low.[4] The conditional probabilities $Cr(\mathsf{pass}|\mathsf{study})$ and $Cr(\mathsf{pass}|\mathsf{party})$ are *act-conditionalized* probabilities. Calculating value expectation with these act-conditioned quantities yields

$$Val_{\text{Jeffrey}}(a) = \sum_k Cr(s_k|a)Val(s_k \wedge a) \tag{2}$$

Using Equation 2, rather than Equation 1, may well recommend in this example that you study instead of party—even though partying dominates studying.[5] Jeffrey's equation is often taken to be the defining equation of EDT.

With Equation 2 in hand, utility maximization can be applied to more cases. But not everyone accepts the equation's verdicts in every case to which it can be applied. Causalists, in particular, believe that Jeffrey's equation overgenerates dependencies.[6] To return to a case mentioned above, it seems that one should not avoid seeking out a therapist on the grounds that x's going a therapist is good evidence that x suffers from anxiety, and thus that *ceteris paribus* $Pr(x \text{ is anxious} \mid x \text{ goes to a therapist}) > Pr(x \text{ is anxious} \mid \neg x \text{ goes to a therapist})$. Even if my credences reflect these probabilities for my own case, what matters is that going to the therapist will not *cause* me to be anxious—and I know this.

In light of such considerations, Gibbard and Harper advance a new utility-maximizing equation, wherein the relevant subjective probability is $Cr(a \boxdot\!\!\rightarrow s_k)$. I will call the object of the agent's credence here the *act-counterfactual* $\ulcorner a \boxdot\!\!\rightarrow s_k \urcorner$, and follow Gibbard & Harper in read-

[3] Savage [23].

[4] For reasons of space I have compressed this introduction. Savage himself, of course, cautioned that the simple Equation 1 was not applicable in dependency cases; see Savage [23] and Joyce [12, Ch. 2] for discussion.

[5] Partying dominates studying in the sense that $Val(\mathsf{party} \wedge \mathsf{pass}) > Val(\mathsf{study} \wedge \mathsf{pass})$—you prefer the former in cases where you pass—and $Val(\mathsf{party} \wedge \mathsf{fail}) > Val(\mathsf{study} \wedge \mathsf{fail})$—you also prefer the former in cases where you fail.

[6] It can undergenerate them as well [10], but for ease of exposition I focus on overgeneration here.

ing it as the subjective probability that *if act a were performed, state s_k would obtain*:

$$\mathcal{U}(a) = \sum_k Cr(a \boxdot\!\!\rightarrow s_k) Val(s_k \wedge a) \tag{3}$$

In a "medical Newcomb" problem like the therapist-anxiety case, although $Cr(\mathsf{axious}|\mathsf{therapist})$ is higher than the baseline $Cr(\mathsf{anxious})$, the agent's credence in the act counterfactual, $Cr(\mathsf{therapist} \boxdot\!\!\rightarrow \mathsf{anxious})$, is *not* higher than the unconditionalized credence $Cr(\mathsf{anxious})$, since counterfactuals track causal influence and going to the therapist (by hypothesis) does not *cause* anxiety. Hence dominance reasoning—the same kind of reasoning that seemed *wrong* in the studying case—may now recommend the pursuit of therapy.

2 Egan's Case

While Egan argues that CDT is the wrong theory of decision, he concedes that it gives the right verdict in Newcomb cases [4, pg. 94]. His counterexample to CDT is—at least at a first glance—very different, and goes as follows:[7]

> **Murder Lesion**. Mary is deliberating about whether to shoot Alfred, a loathsome dictator. She would prefer to shoot him, but only if she will hit him, rather than miss him. Mary has good evidence that a certain kind of brain lesion, which she may or may not have, causes murderous tendencies but also causes shooters to have bad aim. Mary is currently fairly confident that she has good aim, and not very confident she will shoot.

The basic acts in the Murder Lesion case are: $\{\mathsf{shoot}, \neg\mathsf{shoot}\}$, and the basic states are: $\{\mathsf{hit}, \neg\mathsf{hit}\}$. However, Mary's knowledge includes information about causal influence: she has conditional and unconditional subjective probabilities on well-formed formulas like $(\mathsf{shoot} \boxdot\!\!\rightarrow \mathsf{hit})$—*if I were to shoot Alfred, I would hit him.* Egan argues that the set of states relevant to Mary's decision problem should be, not

(Partition 1) $\{\mathsf{hit}, \neg\ \mathsf{hit}\}$

but

(Partition 2) $\{\mathsf{shoot} \boxdot\!\!\rightarrow \mathsf{hit}, \mathsf{shoot} \boxdot\!\!\rightarrow \neg\ \mathsf{hit}\}$.

Using Partition 2, he argues that the act-conditionalized probability relevant in *Murder Lesion* is given by a more complex formula (*), in which both subjunctive and evidential probability are invoked:

$$Cr(a \boxdot\!\!\rightarrow s_k|a) \tag{*}$$

In Mary's case, it is the credence she has on *if I were to shoot, I would hit, given that I shoot.*

I will henceforth call (*) "the Egan credence on s_k in a." Calculating the expected utility of an act a with Egan credences in state-act pairs yields a quantity we can call $\mathcal{U}_{\text{Egan}}(a)$:

$$\mathcal{U}_{\text{Egan}}(a) = \sum_k Cr(a \boxdot\!\!\rightarrow s_k|a) Val(s_k \wedge a) \tag{4}$$

[7] I have shortened the description somewhat to save space; see Egan [4, pg. 97]. For the addition that Alfred is a dictator, see Joyce [13].

Egan argues that, intuitively, Mary should not shoot in *Murder Lesion*. As Mary confronts her decision, $Cr(\mathsf{shoot})$, by description of the case, is low. Therefore, $Cr(\mathsf{shoot} \boxdot\!\!\rightarrow \mathsf{hit})$ is high, since she is relatively confident she does not have the brain lesion. But her Egan credence $Cr((\mathsf{shoot} \boxdot\!\!\rightarrow \mathsf{hit}) \mid \mathsf{shoot})$ is low, since conditional on shoot, she is quite confident she has the lesion, which causes bad aim. Applying Equation 4, we get Egan's favored answer, which is that shooting has a low expected utility.

Egan's view marks an interesting contrast with classical EDT and CDT. He follows CDT in holding, as against EDT, that causal concepts have a direct role to play in decisionmaking. His argument *assumes* that Mary has credences that explicitly invoke causal influence—and, indeed, that she should reason with these credences as she deliberates.

Why, then, does classical CDT go wrong? What Egan's case adds to the dialectical mix is the diachronic norm of conditionalization. Where $Cr^t(\cdot)$ is an agent's credence function at t and $Cr^t_E(\cdot)$ is the credence function she is disposed to adopt upon learning the proposition E, Conditionalization says

> (Conditionalization) For any proposition A, if an agent learns exactly E between times t and t^+,
>
> $$Cr^{t^+}(A) \stackrel{!}{=} Cr^t(A|E)$$

(I use the symbol '$\stackrel{!}{=}$' to indicate that the equality here is intended to be read normatively, rather than descriptively.)

While it is a normative and not a descriptive claim, Conditionalization does in fact characterize much of our ordinary belief revision. As a norm of betting, it is supported by a Dutch Book argument [26]. Importantly for our purposes, the norm apparently entails, as a special case, that one should update by conditionalization on one's own acts. If this is correct, there would seem to be a strong argument in favor of Egan CDT over classical CDT. The force of Conditionalization comes from the thought that, in situations where an agent expects to get more information as time passes, she should regard her future credences as better-informed than her current ones. Egan, in effect, asks: should this not be the case for our future credences *in act counterfactuals*, as well as everything else? Assuming an agent in a decision problem is generally self-aware, she can anticipate what her credence in $(a \boxdot\!\!\rightarrow s_k)$ will be, given that she undertakes a.[8] By Conditionalization, this future credence is just her *current Egan credence* on s_k in a. Reaching for Egan credences in assessing the utility of acts is therefore enlightened thinking: a straightforward application of Jeffrey's appealing claim that a decision-maker should "choose for the person [she] expect[s] to be when [she has] chosen" [11, pg.16].

3 An Inconvenient Proof: The Collapse Lemma

I think this argument has considerable intuitive force. But my presentation has gotten a bit ahead of itself by neglecting to provide a model theory for counterfactual credence. On the usual order of things, the worldly truth-conditions of well-formed formulas ϕ are prior to their probabilities: $Cr(\phi)$ is the amount of probabilistic mass concentrated on the individual *worlds* where ϕ is true. Gibbard & Harper endorse this order of priority for act counterfactuals (op. cit., 127), taking two axioms to characterize the proposition expressed by $\ulcorner a \boxdot\!\!\rightarrow s_k \urcorner$. The first

[8] More carefully, by an agent's being "self-aware", we are assuming that if the agent does a at t^+, she knows it: $Cr^{t^+}(a) = 1$. See, for example, Hare & Hedden [9, pg. 616]: "you are not prone to astounding yourself."

is Modus Ponens. The second is the *Conditional Excluded Middle* (CEM), which embeds the Law of the Excluded Middle ($S \vee \neg S$) under arbitrary counterfactual antecedents:

$$(A \boxdot\!\!\rightarrow S) \vee (A \boxdot\!\!\rightarrow \neg S) \tag{CEM}$$

From these principles, they derive a single characterizing axiom, which they call Consequence 1:

$$A \supset (S \equiv (A \boxdot\!\!\rightarrow S)) \tag{Con 1}$$

With this model-theoretic commitment in hand, we can introduce the Collapse Lemma. It uses Consequence 1 to show that the Egan credence on s in a reduces into the conditional credence in s *given* a. The result is that for any decision problem and any option a, $Val_{\text{Jeffrey}}(a) = \mathcal{U}_{\text{Egan}}(a)$. In other words, Equations 4 and 2 are equivalent.
Proof: by the Ratio Formula,

$$\begin{aligned} Cr_{\text{Egan}}(s \text{ in } a) &= Cr(a \boxdot\!\!\rightarrow s|a) \\ &= Cr((a \boxdot\!\!\rightarrow s) \wedge a)/Cr(a) \end{aligned}$$

By Consequence 1,

$$\begin{aligned} Cr((a \boxdot\!\!\rightarrow s) \wedge a)/Cr(a) &= Cr(s \wedge a)/Cr(a) \\ &= Cr(s|a) \end{aligned}$$

QED.

This is inconvenient for those wishing to press Egan-style counterexamples against CDT: on the leading model-theoretic implementation of the semantic primitives they use, their theory collapses into classical EDT, and so is not a middle-ground position at all. (In particular, it's not clear how Egan can formalize a theory which recommends the dominant option in Newcomb Problems, but also recommends not shooting in *Murder Lesion*.) But the proof is equally inconvenient for CDT. The CDTer must confront, not just Egan's particular counterexamples, but his general *argument*, which brings causal notions into interaction with a norm of learning. It is unsatisfying to rely on Consequence 1 to deprive the argument of force, since the proof may simply indicate that this formulation of CDT is off the mark.

3.1 Strengthening of the Lemma

Where to go from here? An irresistible vantage point on the plausibility of the semantics of counterfactuals is provided by the general framework in Lewis's *Counterfactuals* [17]. According to this familiar, "similarity"-based approach, the counterfactual $\ulcorner A \boxdot\!\!\rightarrow S \urcorner$ is true at a world w just in case S is true at all worlds w' which are both A-worlds and *maximally similar* to w.[9]

As many commentators have pointed out, CEM is often rejected in this framework.[10] Sup-

[9] More precisely: where $v' \leq_v v''$ means that v' is more similar to v than v'' is to v, and where, for any proposition p, $max_{\leq,w}(p) := \{w' : w' \in p \wedge \forall w'' : (w'' \in p) \supset (w' \leq_w w'')\}$, the semantics for the counterfactual is: $u \vDash \ulcorner \phi \boxdot\!\!\rightarrow \psi \urcorner$ iff $\forall u'$ such that $u' \in max_{\leq,u}$, if $u' \vDash \phi$, then $u' \vDash \psi$. For simplicity, I explicitly consider only the case of atomic ϕ and ψ.

[10] The dialectic is a bit complex here: although Gibbard & Harper invoke CEM, their characterizing axiom, Consequence 1, is in fact weaker than CEM in the similarity framework—it is equivalent instead to Strong Centering [17, pg. 132]. The Collapse Argument relies only on Strong Centering, but because without CEM, $Pr(A \boxdot\!\!\rightarrow \cdot)$ is not additive, Gibbard & Harper's appeal the stronger principle is what is plausibly underwriting their commitment to Equation 3 in the first place. For arguments against the weaker (but still strong) Strong Centering principle, see, *inter alia*, Gundersen [8], Leitgeb [14]. For Gibbard & Harper's own reservations about CEM, see op. cit., pg. 128.

pose I am contemplating flipping the coin in my pocket. I know the coin to be fair, so that the outcomes heads and tails would have equal objective probability, or *chance*, of obtaining. It seems strange to hold that the most similar heads-world and the most similar tails-world are not *equally* similar to this world. But CEM allows no ties: it says that either $\ulcorner \mathsf{flipped} \boxdot\!\!\rightarrow \mathsf{heads} \urcorner$ is true and $\ulcorner \mathsf{flipped} \boxdot\!\!\rightarrow \mathsf{tails} \urcorner$ is false, or vice-versa. The principle thus breaks the symmetry between chance-symmetric outcomes in an awkward way.

Lewis [20] rejects Gibbard & Harper's formulation of CDT on these grounds, writing that what he calls "the Chance Objection" is "decisive" against CEM [19, pg. 26]. "Fortunately," he adds, "the needed correction is not far to seek." Lewis's alternative version of CDT is probabilistic, trading determinate propositional consequents for determinate chance consequents, like $\ulcorner Ch(\mathsf{heads}) = .5 \urcorner$. Where Pr is a variable over probability functions and Ch is the objective chance function (see 19, pg. 28), Lewis's new equation is:

$$\mathcal{U}_{\text{Fuzzy}}(a) = \sum_k \sum_{Pr} Cr(a \boxdot\!\!\rightarrow [Ch = Pr]) Pr(s_k) Val(s_k \wedge a) \tag{5}$$

This seems to solve the problem with the coin. Instead of "act-counterfactuals", we might call counterfactuals of the form $\ulcorner a \boxdot\!\!\rightarrow [Ch = Pr] \urcorner$ (for some probability function Pr) "act-to-chance counterfactuals".

Can act-to-chance counterfactuals provide a framework in which to state Egan's *sui generis* CDT, as distinct from both the classical causal and the classical evidential approach? I don't think so. The problem is that while CEM fails in this framework in a narrow sense, the collapse lemma does not rely on the fact that the consequent of $\ulcorner A \boxdot\!\!\rightarrow S \urcorner$ is a proposition. Here's a weaker cousin of CEM explicitly tailored to Equation 5:

$$(A \boxdot\!\!\rightarrow [Ch = Pr]) \vee (A \boxdot\!\!\rightarrow [Ch \neq Pr]) \tag{CEM-Pr}$$

Since different candidate chance functions exclude each other, if A brings about one chance function, it will exclude any other. And that is what (CEM-Pr) says. This version of CEM is actually *more* plausible than its propositional cousin.

Following Equation 5, let $\mathfrak{P}_{\text{Lewis}}(\cdot)$ be the agent's appropriate counterfactual-on-a credence in s_k, broken down into the best estimate of chance:

$$\mathfrak{P}_{\text{Lewis}}(s_k \text{ in } a) = \sum_{Pr} Cr(a \boxdot\!\!\rightarrow [Ch = Pr]) Pr(s_k) \tag{P1}$$

Once again, we construct the corresponding Egan credence:

$$\mathfrak{P}_{\text{Egan}}(s_k \text{ in } a) = \sum_{Pr} Cr(a \boxdot\!\!\rightarrow [Ch = Pr] | a) Pr(s_k) \tag{P2}$$

By a similar proof to the one above, it can be shown that $\mathfrak{P}_{\text{Egan}}(s \text{ in } a) = \mathfrak{P}(s|a)$.[11]

[11] As before:

$$\begin{aligned}
\mathfrak{P}_{\text{Egan}}(s \text{ in } a) &= \sum_{Pr} Cr(a \boxdot\!\!\rightarrow [Ch = Pr] | a) Pr(s) \\
&= \sum_{Pr} [Cr((a \boxdot\!\!\rightarrow [Ch = Pr]) \wedge a) / Cr(a)] Pr(s) \\
&= \sum_{Pr} [Cr([Ch = Pr] \wedge a) / Cr(a)] Pr(s) \\
&= \sum_{Pr} Cr([Ch = Pr] | a) Pr(s)
\end{aligned}$$

QED.

We should also note that while (CEM-Pr) looks reasonable, it isn't immediately clear how to model chance consequents in the Lewis similarity framework. In order for it to make sense to say that, in all the most w-similar worlds where A is true, the chance function is (some particular) Pr, one would need to make sense of chances holding at individual worlds, rather than being irreducible features of how probability is spread across the *space* of worlds. Lewis, in his classic statement of chance, is indeed comfortable with worldbound chances, but it is unusual in the typical Bayesian framework.[12]

3.2 Concluding Thoughts

To conclude, I want to return briefly to the question of how a Causalist working in the style of Gibbard & Harper can reply to Egan's conceptual argument about anticipated causal credences.

Our initial puzzle was that, conceptually, it seemed like there were three things: (i) the probability of a state, given an act; (ii) the probability that if the act *were* performed, the state *would* result; and (iii) the probability one would have in that same counterfactual, if one learned the act was actually performed.[13] But it then turned out that (i) and (iii) couldn't be distinguished on a first-pass semantics for the counterfactual. As we saw, the argument initially looks implausible because of its reliance on CEM. But when CEM is weakened in a plausible way, the collapse actually recurs—apparently, with renewed strength.

Therefore, one response to Egan's counterexample is just to rely on the lemma to suggest that the appearance of there being three things, rather than two, was mistaken. The argument from future credence in counterfactuals was just a dressed-up version of the same appeal Causalists learned to reject in Newcomb problems. Moreover, the causal decision theorist can provide a complete model theory compatible with act-to-chance counterfactual view. It is just the Lewisian "closeness" theory, with chances construed as world-bound.

A very different response—one that I favor, but lack the space to fully defend here—also gains strength from the move to chance counterfactuals. But it takes a global, rather than local, perspective on those chance counterfactuals.

Recall that Conditionalization is a norm for getting from one credal state to another, when something is learned. Egan's conceptual argument relied on the strength of this norm. But it is well-known that some sentences in natural language cannot express worldly truth-conditions consistently with this role. Acceptance of such sentences seems to update one's credal state, but it cannot do so by conditionalization.

For example, suppose you learn that *might* A (where 'might' is read epistemically). What should the effect on your credences be? It is hard to believe that the following is *not* an applicable credal norm:

> (Norm-'might') For any proposition A, if the agent learns exactly $\ulcorner$might $A\urcorner$ between times t and t^+,
>
> $$Cr^{t^+}(A) \overset{!}{>} 0.$$

Note here that pooling does not commute with conditionalization, so $\sum_{Pr} Cr([Ch = Pr]|a)Pr(s) \neq \sum_{Pr} Cr([Ch = Pr])Pr(s|a)$ (5, 25, 15).

[12] See e.g. Lewis [16, pg. 269]: "The term 'the chance, at t, of A's holding' is a nonrigid designator...it designates different numbers at different worlds." Once again, coins may be helpful in cashing out such a picture. For example, in any world where I flip Coin 1—regardless of how it lands—there is a certain chance of heads. This chance is different in any world where I flipped the differently weighted Coin 2 instead—even both absolute outcomes, heads and tails, happen in each type of world.

[13] For example, (i) the probability that I am anxious, given that I see a therapist; (ii) the probability that *if I were to see a therapist, it would make me anxious*, and (iii) the probability of (ii), given that I just learned that I *do* see a therapist.

A miniature, two-line "triviality proof" (for example, in the style of 22) can show that (Norm-'might') conflicts with the norm of Conditionalization.[14] Other epistemic expressions ('must', 'probably') and, of course, indicative conditionals (via 18) have the same character.

There is no reason the chance counterfactuals of Equation 5 should not be the same way. On this view, accepting a chance counterfactual puts global constraints on an agent's credences directly—not indirectly, via Conditionalization. Lewis's discussion of probabilistic counterfactual *chances* makes it especially clear how this would proceed: it would go via the Principal Principle [16], the norm that one should conform one's credences to objective chance.

For example, for full belief that one's acts bring about a particular set of chances, we could write the norm as follows:

> (Norm1-$\Box\!\!\rightarrow$) For any state S, if $Cr^t(a \mathbin{\Box\!\!\rightarrow} [Ch = P]) = 1$ and the agent performs act a between times t and t^+,
>
> $$Cr^{t^+}(s) \overset{!}{=} P(s)$$

This norm overrides Conditionalization, just as (Norm-'might') does.[15]

This begins to give the CDTer a reply to Egan's conceptual argument. The Causalist's reply should target the underlying appeal to Conditionalization that drives the argument. As other epistemic vocabulary shows, Conditionalization is not a plausible guide to belief revision for *every* well-formed formula. Such formulas are not, however, "lawless"; they simply are governed by more specific normative constraints. In the case of act-to-chance counterfactuals, one could, for example, plausibly substitute (Norm1-$\Box\!\!\rightarrow$) instead.[16]

The claim that counterfactuals generate global constraints on credences, and thus cannot express ordinary, "intersective" propositions, has been made before, on the basis of a somewhat different norm than the one I've advanced here.[17] My concern in this short piece has been

---

[14] Suppose for reductio that (Norm-'might') is compatible with Conditionalization. Then there is some proposition p (= the one expressed by $\ulcorner$might $A\urcorner$) such that, for any probabilistic credal state $Cr(\cdot)$,

$$Cr(A|p) \overset{!}{>} 0$$

Clearly *this* cannot be so, for Cr may rule out A on independent grounds. Begin with an arbitrary probability function $Cr(\cdot)$, and let $Cr'(\cdot)$ be the result of updating Cr with $\neg A$ (viz., so that $Cr'(\cdot) = Cr_{\neg A}(\cdot)$). Now, by Conditionalization,

$$Cr'(A|p) \overset{!}{=} Cr(A|\neg A \wedge p) = 0$$

...contradicting (Norm-'might'). See Russell & Hawthorne [22, §3].

[15] For cases of uncertainty—where your confidence is divided between several such Pr—the norm, following Lewis's formulation of $\mathcal{U}_{\text{fuzzy}}$, is:

> (Norm2-$\Box\!\!\rightarrow$) For any state S, if the agent performs act a between times t and t^+,
>
> $$Cr^{t^+}(s) \overset{!}{=} \sum_{Pr} Cr^t(a \mathbin{\Box\!\!\rightarrow} [Ch = P])P(s)$$

(Norm2-$\Box\!\!\rightarrow$) entails (Norm1-$\Box\!\!\rightarrow$) as a special case.

[16] Here is a sketch of a mini-Triviality result for (Norm1-$\Box\!\!\rightarrow$). Suppose the agent at t has full belief (credence 1) in $a \mathbin{\Box\!\!\rightarrow} [Ch = Pr']$ for some particular probability function Pr' such that $Pr'(s) = .1$. Moreover, $Cr^t(s) = 0$. The agent performs act a. By the norm, it should be the case that $Cr^{t^+}(s) = .1$. But this cannot be the result of conditionalizing $Cr^t(\cdot)$ on any proposition. The important contrast here is with Lewis [18]'s proof that—given CEM— $Pr(C \mathbin{\Box\!\!\rightarrow} X) = Pr^C(X) = \sum_w Pr(w)Pr(X|w_C)$, where w_C is the unique w-closest C-world (whose existence is guaranteed by CEM).

[17] See especially the discussion in Chapter 6 of Joyce [12], as well as Williams [27] and Briggs [3].

decision theoretic, and thus independent of at least *many* features of the behavior of counterfactuals in natural language. I *do* think it is difficult to defang the argument underlying Egan's counterexamples. But perhaps this is a start.

References

[1] Arntzenius, Frank (2008). "No Regrets, or: Edith Piaf Revamps Decision Theory." *Erkenntnis*, 68(2): pp. 277–297.

[2] Briggs, R.A. (2010). "Decision-Theoretic Paradoxes as Voting Paradoxes." *Philosophical Review*, 119(1): pp. 1–30.

[3] Briggs, R.A. (2017). "Two Interpretations of the Ramsey Test." In Beebee, Hitchcock, and Price (eds.) *Making a Difference: Essays on the Philosophy of Causation*, Oxford University Press.

[4] Egan, Andy (2007). "Some Counterexamples to Causal Decision Theory." *The Philosophical Review*, 116(1): pp. 93–114.

[5] Genest, Christian, and James Zidek (1986). "Combining Probability Distributions: A Critique and an Annotated Bibliography." *Statistical Science*, 1: pp. 114–148.

[6] Gibbard, Allan, and William Harper (1978). "Counterfactuals and Two Kinds of Expected Utility." In Hooker, Leach, and McClennen (eds.) *Foundations and Applications of Decision Theory, Vol 1*, Dordrecht: D. Reidel.

[7] Greaves, Hilary (2013). "Epistemic Decision Theory." *Mind*, 122(488): pp. 915–952.

[8] Gundersen, Lars (2004). "Outline of a New Semantics for Counterfactuals." *Pacific Philosophical Quarterly*, 85(1): pp. 1–20.

[9] Hare, Caspar, and Brian Hedden (2015). "Self-Reinforcing and Self-Frustrating Decisions." *Noûs*, 50(3): pp. 604–628.

[10] Hesslow, Grant (1976). "Discussion: Two notes on the probabilistic approach to causality." *Philosophy of Science*, 43(2): pp. 290–292.

[11] Jeffrey, Richard (1983). *The Logic of Decision.* University of Chicago Press.

[12] Joyce, James (1999). *The Foundations of Causal Decision Theory.* Cambridge University Press.

[13] Joyce, James (2012). "Regret and Instability in Causal Decision Theory." *Synthese*, 187: pp. 123–145.

[14] Leitgeb, Hannes (2012). "A Probabilistic Semantics for Counterfactuals." *Review of Symbolic Logic.*

[15] Leitgeb, Hannes (2016). "Imaging All The People." *Episteme*, pp. 1–17.

[16] Lewis, David (1971). "A Subjectivist's Guide to Objective Chance." *Studies in Inductive Logic and Probability 2.*

[17] Lewis, David (1973). *Counterfactuals.* Oxford: Blackwell.

[18] Lewis, David (1976). "Probabilities of Conditionals and Conditional Probabilities." *Philosophical Review*, 85.

[19] Lewis, David (1981). "Causal Decision Theory." *Australasian Journal of Philosophy*, 59(1): pp. 5–30.

[20] Lewis, David (1981). "Why Ain'cha Rich?" *Noûs*, 15: pp. 377–80.

[21] Richter, Reed (1984). "Rationality Revisited." *Australasian Journal of Philosophy*, 62(4): pp. 392–403.

[22] Russell, Jeffrey, and John Hawthorne (2016). "General Dynamic Triviality Theorems." *Philosophical Review*, 125(3): pp. 307–339.

[23] Savage, Leonard (1972). *The Foundations of Statistics*. Dover.

[24] Stalnaker, Robert (1981). "A Defense of the Conditional Excluded Middle." In Harper, Stalnaker, and Pearce (eds.) *Ifs: Conditionals, Belief, Decision, Chance, and Time*, D. Reidel.

[25] Steele, Katie (2012). "Testimony as Evidence: More Problems for Linear Pooling." *Journal of Philosophical Logic*, 41: pp. 983–999.

[26] Teller, Paul (1973). "Conditionalization and Observation." *Synthese*, 26: pp. 218–258.

[27] Williams, J.R.G. (2012). "Counterfactual Triviality: A Lewis-Impossibility Argument for Counterfactuals." *Philosophy and Phenomenological Research*, LXXXV(3): pp. 648–670.

But, scalar implicatures and covert quotation operators *

Yael Greenberg

Bar-Ilan University, Ramat Gan, Israel
yaelgree@gmail.com

Abstract

This paper deals with a cross linguistically productive, yet puzzling construction which we call *x but (really) X* (e.g. *John is always but (really) ALWAYS late*), which surprisingly combines contrast and strengthening. We examine, but reject, an initial analysis where the conjuncts of ***but*** in this construction are domain-based or degree-based scalar alternatives and where the second, semantically stronger conjunct rejects a scalar implicature of the first. We then develop a revised analysis where ***but*** is under the scope of a covert quotation operator (as in 'mixed quotations'). The analysis captures the 'metalinguistic' flavor of *x but X* and the contribution of the 'contrastive' semantics of ***but*** to the strengthening effect of the whole construction, and is shown to avoid over-generation risks.

0 Introduction

This paper deals with a construction which we will call *x but X* (following the terminology in Greenberg (2014), Shitrit (2015)). Here are two examples in Hebrew, where this construction is very productive and common:

(1) a. *kulam, aval (mamaS) KULAM higi'u*
everybody but really everybody arrived

"Everybody but really EVERYBODY arrived" (="Absolutely everyone arrived")

b. *Dani gavoha, aval (mamaS) GAVOHA*
Danny tall but really TALL

"Danny is tall, but really TALL" (= "John is very tall")

The sentences in (1) are characterized by the presence of a contrast particle *aval* ('but') - conjoining two expressions with the same lexical content, e.g. the two identical universal quantifiers *kulam* ('everybody') or the two identical gradable adjective *gavoha* ('tall'), with an optional adverb like *mamaS* ('really'). The second conjunct (henceforth *x2*) is accented and seems to be interpreted as 'stronger' or intensified relative to first conjunct (henceforth *x1*). In (1b), for example, *x2* is understood as "Absolutely everyone, with no exceptions" and in (1b) it is understood as 'definitely / very tall'. The overall effect of the construction is strengthening, as seen in the glosses.

This 'strengthening' effect of *x but X* seems rather clear, but the derivation of this effect, and in particular, the contribution of the contrast particle (*aval* - 'but') to the compositional interpretation of the construction is not clear at all, at least not immediately. Why is a contrast particle used in a 'strengthening' construction? Isn't there a clash between 'contrast' and 'strengthening' to start with?

*Thanks to Mira Ariel, Elitzur Bar Asher Siegal, Ariel Cohen, Luka Crni, Danny Fox, Julie Goncharov, Andreas Haida, Yosi Grodzinski, Roni Katzir, Sven Lauer, Emar Maier, Lena Miashkur, Susan Rothstein, Galit Sassoon and the audience at TAU, the LLCC at HUJI, ESSLLI 2016 and IATL 2016 for constructive discussions and comments. Work on this research is supported by ISF grant 1655/16 to Yael Greenberg.

One could think that *aval* is not used in its original, contrastive meaning in (1), or perhaps that *x but X* has an idiomatic use, unique to Hebrew. But this does not seem to be the case, as many other language use a contrast particle to achieve similar strengthening effects in parallel *x but X* constructions, though their means for strengthening *x2* vary. While in Hebrew mere accentuation of *x2* is enough, and an intensifying adverb is optional, in French / English / German the construction is to a large extent natural only in the presence of the intensifying adverbial (*All, but really ALL*[1]*....* / *tout mais vraiment tout....* / *Alles aber wirklich alles*[2]*...*). Spanish, on the other hand, can use reduplication of *x*2 (*todo, pero todo todo...*), and, as Hoeksema (2007) reports, Dutch can use an additive particle *(ook)* with *x2* as in *Nooit maar dan ook nooit weer* (translated as "Never, absolutely never again").[3]

In what follows we will abstract away from these different strategies and concentrate on the core features of *x but X*, and in particular on the presence of the contrast particles (e.g. *aval, pero, mais, aber, maar, but,* henceforth ***but***) in it and their contribution to the overall effect of this 'strengthening' construction.

To do that we will start in section 1 by reviewing an initial analysis of *x but X* in Hebrew, developed in Greenberg (2014), Shitrit (2015), which uses a standard counter-expectational semantics for *but* (as in e.g. Winter and Rimon (1994), Toosarvandani (2014)) and analyzes the conjuncts in this constructions as domain-based or degree-based scalar alternatives. The analysis then derives the strengthening effect in *x but X* from assuming that the stronger conjunct, *x2* rejects a scalar implicature of the weaker *x1*. In section 2, however, we show that this initial analysis cannot hold, since in 'normal' *a but b* sentences *but* is systematically banned from being used to reject scalar implicatures when its conjuncts are scalar alternatives (cf. Winterstein (2013)). In section 3, then, we develop a revised analysis of the *x but X* construction, according to which it expresses metalinguistic contrast, where intuitively what is taken into account is not our knowledge of *x1* and *x2* (as in regular *a but b* sentences) but rather the knowledge that the speaker used *x1* and used *x2* to convey a single meaning. The intuition is formally captured by placing *but* (with its usual counter-expectational semantics) under an independently argued for, though covert quotation operator (cf. von Fintel (2004), Geurts and Maier (2003), Maier (2014, 2017)), and it is shown to avoid over-generation risks. Section 4 concludes and examines some general implications of the proposal and questions left open for future research.

1 An initial analysis of 'x but X'

Shitrit (2015) proposes an analysis of *x but X* which has three ingredients. First, she adopts Greenberg (2014) suggestion that the conjuncts in sentences like (1) are scalar alternatives (cf. Fox and Katzir (2011), Katzir (2014)), which are domain-based or degree-based.[4] For example,

[1]Though examples with no accompanying adverbial are attested as well, as in (i) and (ii) (found on a Google search):
(i) *Always, but ALWAYS back up your SD card and files regularly*
(ii) *No one, but no one does childhood nightmares better than Roald Dahl*

[2]Thanks to Sven Lauer (p.c.) for this German data.

[3]Hoeksema (2007) discusses this construction in the context of parasitic licensing of NPIs, and calls it 'emphatic reduplication'. However, since this title can potentially cover other constructions (e.g. the 'salad salad' cases, discussed in Ghomeshi et al. (2004)), and does not reflect the contribution of the contrast particles, which is the focus of this paper, I keep the title *x but X*.

[4]See Kadmon and Landman (1993) on domain widening with *any* and Chierchia (2013) on degree-based and domain-based alternatives with covert *E(ven)* and *O(only)* (involved in the semantics of some NPIs). See also Greenberg (2016b), Greenberg and Orenstein (2016) on overt *even*-like and *only*-like operators over degree-based

in (1a) *x1* and *x2* are domain-based scalar alternatives, derived by assigning the covert domain variable D a 'default' (narrow) value and a 'wide' value, as seen in (2):[5]

(2) $[\forall x \in D_{default}$ arrived(x)] but $[\forall x \in D_{wide}$ arrived (x)] (where $D_{default} \subset D_{wide}$)

In (1b) the conjuncts are 'degree-based' alternatives, derived by assigning the covert standard variable in the positive form (cf. Kennedy and McNally (2005)) a default value and a 'high' value, as in (3):

(3) $[\exists d\ d \geq standard_{default,tall} \wedge tall(s,d)]$ but $[\exists d\ d \geq standard_{high,tall} \wedge tall(s,d)]$
where $standard_{high,tall} > standard_{default,tall}$

Second, Shitrit suggests, the contrast particle *but* in '*x but X*' has a regular counter-expectational semantics (cf. Winter and Rimon (1994), Toosarvandani (2014)), intuitively characterized in (4) and illustrated in (5):[6]

(4) *a but b* asserts the conjunction of *a* and *b* and presupposes that there is a cancellable implication of *a,* ('*r*') that *b* rejects (i.e. implies or entails its negation)

(5) a. *It was raining but we remained dry*
b. *It was raining but we took an umbrella*

Given (4), then, (5a) and (5b) assert that both conjuncts are true and presuppose that there is a proposition *r* (e.g. *We got wet*) which is implied by *a* (*It was raining*) and is rejected by *b*, i.e. its negation is entailed by *b* (in the case of *We remained dry* in (5a)), or implied by it (in the case of *We took an umbrella* in (5b)).

The third, and most important ingredient in Shitrit's analysis is the suggestion that in the case of *x but X,* the proposition *r* that *x1* implies and *x2* rejects is in fact a scalar implicature of *x1.* For example, the scalar implicatures which *x1* imply in (2) and (3) are *Everyone only in the default domain arrived* and *John is only at least as tall as the default standard*, respectively. Indeed, the stronger conjuncts, *x2*, reject these scalar implicatures, so the 'counter-expectational' presupposition of *but* is satisfied.

The suggestion, then, has the advantage of capturing the contribution of ***but*** to the strengthening effect of *x but X* and deriving it from the standard semantics of ***but*** and some minimal assumptions regarding the semantics of its conjuncts, and a general mechanism of scalar implicatures, with no further stipulative steps.

2 A challenge to the initial analysis

Despite the advantages of the initial analysis above, there is also a serious problem it faces. When we try to use regular *a but b* sentences to express rejection of scalar implicatures of *a*

and domain-based alternatives.

[5] Shitrit analyzes the construction as always conjoining two propositions, and as involving ellipsis as in (i):
(i) *Everyone ~~arrived~~ but (really) EVERYONE arrived*
Another option is to assume a cross-categorical analysis of *but,* similarly to that of *and* (as in Partee and Rooth (2008) and subsequent work). We will not try to develop these directions further here, and will henceforth use the term 'conjuncts' loosely, referring to the two arguments of *but*, whatever their type is.

[6] We ignore here the attempts to unify all uses of *but* (see, e.g. Jasinskaja (2012), Jasinskaja and Zeevat (2008), Umbach (2005), Toosarvandani (2014),Winterstein (2013) etc).

by *b*, the result is strikingly infelicitous. This is illustrated, for example, in the infelicity of sentences like (6) (noted by Winterstein (2013)), where *b* (*All students arrived*) rejects the well known scalar implicature of *a* (*Some students arrived*), namely the implicature that not all students arrived:

(6) #Some students arrived but all did

This picture is not limited to quantifiers but seems much more general. We get the same kind of infelicity in (7)-(10) as well:

(7) #I like John, but I love him
(8) #John is good but superb at math
(9) #You can but must leave now
(10) #This is possible but necessary

But, then, is systematically infelicitous when it appears with what we will call ***weak but strong*** constructions, whose second conjunct, which is a stronger scalar alternative to the first, rejects a scalar implicature of the first. An obvious question, of course, is what the reason is for this general ban on *weak but strong* constructions. This is not a question we will focus on in this paper.[7] Crucially, though, whatever the reason is, under the initial analysis above, this reason should equally apply to *x but X* sentences, as in (1), as well, since here too the second conjuncts (*x2*) are stronger scalar alternative of the first (*x1*), and here too *x1* implicates a scalar implicature that *x2* rejects. I.e. the initial analysis analyzes *x but* X sentences as *weak but strong* as well, and thus wrongly predicts it to be as infelicitous as (6)-(10).

Since this is not the case − *x but X*is productive and felicitous − we suggest that the initial analysis should be revised.

3 A revised analysis of *x but X*: *but* is under scope of a metalinguistic, quotation operator

3.1 The intuition

The revised analysis is inspired by an intuitive observation made in Greenberg (2014), Shitrit (2015), namely that the *x but X* construction has a 'metalinguistic' flavor. For example that (1a) can be paraphrased as *Everyone, and (when I say "everyone") I mean that EVERYONE, arrived!* Given this intuitive observation we propose that unlike what happens in normal *a but b* sentences like (5) above, what we consider with *x but X* is not the knowledge of *a (x1)* and *b* (*x*2), but rather the knowledge that the speaker uttered *a*, and that she uttered *b,* together with the fact that she used the two utterances to convey a single meaning. This is seen in the following paraphrases of (1a):

(11) a. By uttering *EVERYONE* I meant to reject the implication that what I meant by uttering *everyone* was something like "*Only everyone in the default, narrow domain,*

[7] See Winterstein (2013) for an argumentative explanation on the infelicity of (6), modeled using conditional probability (cf. Anscombre and Ducrot (1984), Merin (1999)). See also Greenberg (2016a) for a QUD-based + redundancy explanation (cf. Toosarvandani (2014), QUD-based semantics of *but* and Fox (2007), Shitrit (2015), Mayr and Romoli (2016) on redundancy constraints).

D_{default}." I actually used BOTH forms to convey the same thing, namely "Everyone in the wide domain, D_{wide}"

b. You might infer from my utterance of *everyone* (applied to *arrive*) that I meant that only everyone in the default domain, D_{default}, arrived, but this inference should be rejected: I meant the same thing as I mean when uttering *EVERYONE* (applied to *arrived*), i.e. that everyone in a wide domain, D_{wide}, arrived, i.e. that everyone, with no exceptions, arrived

As we will show below, the result of analyzing (1a) this way is that *x but X* is not a *weak but strong* construction, since its conjuncts are not scalar alternatives anymore. If this is so then whatever the problem is for *weak but strong* constructions (e.g. *#Some but all students arrived*), it will not be present with *x but X* sentences, hence accounting for their felicity.

The challenges we face now, then, are (a) how to capture the intuition in (11) a precise way, (b) how to show that, given this proposal, *x but X* sentences indeed escape the general ban on *but,* since its conjuncts are not scalar alternatives and (c) how not to over-generate metalinguistic uses of (6)-(10) above.

3.2 Quotation operators

To capture the intuition in (11) what we need is an explicit and transparent mechanism capturing the shift from utterances (of phonological strings) to meanings. Luckily, such a mechanism was already independently developed, to capture **mixed quotations** I.e. quotations which are syntactically and semantically incorporated into the sentence, but which still maintain the information that they are mentioned. Two examples are seen in (12), and we follow theories like von Fintel (2004), Winter and Rimon (1994), Maier (2014, 2017) who take them to be interpreted as in (13a,b):

(12) a. Quine said that quotation "has a certain anomalous feature" (Davidson (1979))

b. Bush said that the enemy "misunderestimated me" (Maier 2008)

(13) a. Quine said that quotation has what he referred to as a certain anomalous feature

b. Bush said that the enemy did what he referred to as misunderestimated me

Given these theories in a sentences like (12b) the quotation of the phonological string *misunderestimated me* is interpreted as the definite description in (14):

(14) ι A [refer (s, ***"misunderestimated me"***, A)]

In prose: The unique semantic object A (here a property), such that the source s (here Bush) referred to A by using the phonological string *misunderestimated me*. Then (12b) ends up with the intuitive interpretation in (15):

(15) Bush said that the unique property A that he (Bush) referred to by using the phonological string *misunderestimated me* is true of the enemy

Following the above mentioned theories we take ι in (14) to trigger a uniqueness presupposition, i.e. the presupposition that there indeed exists a unique property *A* that *s* (Bush, in this case) referred to by using the phonological string *misunderestimated me*[8].

[8]For simplicity we follow here von Fintel (2004) who takes the definite to pose a definedness condition. But

3.3 Analysis

We now propose that *but* in *x but X* is under the scope of a quotation operator, which is covert (i.e. not marked by quotation marks). (1a) (*Everyone, but (really) EVERYONE arrived*), for example, is analyzed as in (16):

(16) ι $GQ_{<<et>,t>}$ [[refer (s, 'everyone', GQ)] **but** [refer (s, 'EVERYONE', GQ)]] ($arrived_{<e,t>}$)

We continue to assume that *but* has its usual counter-expectational semantics in (1a) (see again (4) above). Given (16), then, (1a) asserts that the two conjuncts of *but* are true, i.e. that the unique GQ, such that the speaker referred to this GQ by uttering *everyone*, and she referred to this same GQ by uttering *EVERYONE,* is true of the property *arrived*.

In addition, there are two presupposition triggering operators here: ***but,*** triggering its counter-expectational presupposition, and the definite in the quotation operator, i.e. ι, triggering a uniqueness presupposition. Importantly, in the case of the felicitous (1a) we can now show that both presuppositions are indeed met.

First, for the counter-expectational presupposition of ***but*** to be met in (16), we need to show that there is a proposition r which is implied by *x1* and rejected by *x2*. We suggest that in our case r is the identity proposition in (17) (where *exh* is used as in Fox (2007), Chierchia et al. (2011)):

(17) r for (1a): GQ= λP. exh$\forall$x $\in$ $D_{default}$ P(x)

Indeed, learning that the speaker uttered *everyone* (and not *EVERYONE*) to refer to the unique GQ, we may draw the implication that she used *everyone* to refer to the exhaustified quantifier "Only everyone in the narrow domain", and learning she uttered *EVERYONE* to refer to the unique GQ, can be indeed taken to reject this implication.

Second, the presupposition of ι, requiring that there is a unique GQ that the speaker actually referred to by uttering *everyone*, and referred to by uttering *EVERYONE,* seems to be met in (1a) as well. We suggest that this unique GQ that speaker refers to is the one in (18):

(18) λP. $\forall$x$\in$$D_{wide}$ P(x)

That is, the idea is that the speaker actually referred to the same quantifier, namely "Everyone in the wide domain", both when uttering *everyone* and when uttering *EVERYONE*. This captures the intuitive paraphrase of (1a) in (11) above as: "and when I say 'everyone' I mean the same thing that I mean when I say 'EVERYONE'. I.e. eventually I mean that everyone in the wide domain arrive".

In addition, and unlike the initial analysis of *x but X* reviewed in section 2, the fact that such sentences are not infelicitous as the ones in (6)-(10) is not problematic anymore. This is because in this revised analysis the conjuncts are not scalar alternatives, so *x but X* is actually not a *weak but strong* construction. For example *refer (s, "everyone", GQ)* is not a scalar alternative to (it does not entail) *refer (s, "EVERYONE", GQ)*. Hence, the general ban on *weak but strong* constructions is not violated with *x but X* sentences, explaining why they are not infelicitous.

Finally, the analysis does not over-generate metalinguistic readings of infelicitous sentences like (6)-(10). To see why this is the case, consider two scopal options for analyzing sentences

see Winter and Rimon (1994), Maier (2014, 2017) who develop a dynamic, DRT-based analysis for coping with cases where the presupposition is cancelled or accommodated.

like (6) (*#Someone but everyone arrived*), namely putting *but* under a quotation operator (as we did for *Everyone, but EVERYONE*) or putting *but* above the quotation operator. Crucially, with both options we end up with a problematic result.

Consider the first option for analyzing (6), where *but* is under the quotation operator, as in (19):

(19) $\iota GQ_{<<et>,t>}$ [[(refer (s, "someone", GQ)] but [refer (s, "everyone", GQ)] (arrived$_{<<et>,t>}$)

(19) asserts that the unique Generalized Quantifier such that the speaker referred to this GQ by uttering *someone* and the speaker referred to this same GQ by uttering *everyone*, is true of the property denoted by *arrive*. We suggest that in this case the uniqueness presupposition triggered by the definite fails. In particular, whereas with (1a), analyzed as (16) (*Everyone but EVERYONE arrived*) we can indeed assume that there is a unique quantifier that that the speaker who utters both forms is eventually referring to, this does not seem to be the case with (6), analyzed as (19). Here there seems to be no reasonable unique GQ that the speaker referred to this GQ by uttering *someone*, and referred to this same GQ by uttering *everyone.*

If we try the second option and put *but* above the quotation operator the result is (20):[9]

(20) $\iota GQ_{<<et>,t>}$ [[refer (s, "someone", GQ) but $\iota GQ_{<<et>,t>}$ [refer (s, "everyone", GQ)]] (arrived$_{<e,t>}$)

(20) asserts that the unique Generalized Quantifier such that the speaker referred to this GQ by uttering *someone,* and the unique Generalized Quantifier such that the speaker referred to this GQ by uttering *everyone,* are both true of the property denoted by *arrive.*

Here, unlike (19), there can be two distinct GQ that the speaker refers to. However, this interpretation fails as well. We assume that the second conjunct in (20) - the unique Generalized Quantifier such that the speaker referred to this GQ by uttering *everyone* - denotes the GQ in (21) (we abstract away from domains now):

(21) $\lambda P.\forall x\ P(x)$

The first conjunct- the unique Generalized Quantifier such that the speaker referred to this GQ by uttering *someone* - can have one of the two denotations in (22):

(22) a. $\lambda P.\exists x\ P(x)$ (i.e. at least someone)

b. $\lambda P.$ exh $\exists x\ P(x)$ (i.e. only someone)

Crucially, in either of these two interpretations, (20) comes out infelicitous: If (22a) is chosen, we end up, once again, with an infelicitous 'weak but strong' case, since $\lambda P.\exists x\ P(x)$ (*at least one)* and $\lambda P.\forall x\ P(x)$ (*everyone*) are scalar alternatives where the second is stronger than the first. If, on the other hand, (22b) is chosen, we end up with a contradiction in the assertion, since *only someone arrived and everyone arrived* is contradictory.

Thus, unlike (1a) (*Everyone but EVERYONE arrived*), sentences like (6) (*Someone but everyone arrived*) cannot be given a felicitous metalinguistic interpretation. This holds for all cases in (7)-(10) as well. We are left with a standard *weak but strong* interpretation of such sentences, which, as we noted above, is generally banned, hence their infelicity.

[9] Here we will have to take *but* to denote cross categorical conjunction. See footnote #4 above.

4 Conclusion and directions for future research

The challenge we dealt with in this paper is how to give a compositional interpretation of a cross linguistically productive, yet puzzling construction - *x but (really) X* (e.g. *John is always but (really) ALWAYS late)* - with a surprising interaction of contrast and strengthening. We examined, but rejected, an initial analysis where the conjuncts of *but* in *x but X* are scalar (domain-based or degree-based) alternatives and where the second and semantically stronger conjunct rejects a scalar implicature of the first. We then developed a revised analysis of the construction, based on the idea that this construction expresses metalinguistic contrast, and captured it by placing ***but*** under the scope of a covert quotation operator (independently developed to capture 'mixed quotations'). We showed that this analysis indeed captures the contribution of the 'contrastive' semantics of ***but*** to the strengthening effect of the whole construction, that it does not run into the problem with the initial analysis and that it does not over-generate.

Besides explaining the puzzling interaction between contrast and strengthening in this construction, the analysis provides support for the linguistic relevance of metalinguistic operations, and more specifically, for the contribution of quotation operators to the compositional interpretation of a wider set of sentences than have been initially considered in the literature on quotations. More research is needed to examine the behavior of the construction with respect to other features of mixed quotations such as indexical and language shifts, cancellation and / accommodation of the uniqueness presupposition of the quotation operator, 'unquotation', etc. (cf. Maier (2014, 2017), Shan (2010) and others). An interesting question is why speakers use the *x but X* construction to start with, instead of a simple intensified form (e.g. *John is always, but ALWAYS late*, instead of the simpler *John is ALWAYS late*).[10] The productivity of the construction, though, seems to indicate that it does serve a certain purpose, perhaps that of indicating a more total exclusion of exceptions and / or self correction.

An important puzzle for future research concerns Hoeksema (2007) and Shitrit (2015) obesrvations (about Dutch and Hebrew, respectively) that *x but X* is much more common with universal quantifiers (and gradable predicates) than with existential quantifiers. In English too there are hardly any attested occurrences of *Sometimes but really sometimes...* as opposed to *Always but really always....* We note, though, that *Sometimes, but only sometimes...,* with an explicit exclusive, instead of an intensifying adverb (and / or accentuation) is much better and much more commonly attested. We leave the examination of these patterns to future research.

Another question is whether the above analysis of *x but X* can cope with cases like (23):

(23) a. Obama, but (really) OBAMA called me (cf. Shitrit 2015)

b. The night guard must close both, but (really) BOTH doors[11]

On the surface these cases seem to differ from those in (1) since they do not seem to involve any scale, domains or degrees. But perhaps they can be analyzed as potentially scalar after all. The intuitive effect in (23a) can be that the speaker really means that Obama, and not, e.g. one of his secretaries called, which can be paraphrases with an intensified 'self' ("Obama, but (really) Obama himself, called me"). This can be perhaps modeled by assuming the operation of a covert *even* operator in the construction (cf. Charnavel (2015)), and then, indirectly, mapping the conjuncts to a scale of e.g. importance / prestige (cf. Greenberg (2015, 2017) for a semantics of *even* which encodes such mappings). In the case of (23b) we

[10]Thanks to Emar Maier (p.c.) for this point.

[11] Thanks to Andreas Heida (p.c.) for this example.

seem to get a higher commitment effect, which can be paraphrased as "The night guard must close both door, and I am completely serious about this! No joking!". This can be naturally uttered in a situation where the night guard many times forgets basic things, or interprets our instructions lightly. A potential direction for accounting for such cases is manipulate degrees of commitments or credence the speaker has towards *p,* as in a 'gradable' modeling of the *ASSERT* speech act operator, developed in e.g. Wolf (2015), Greenberg and Wolf (2017). More research, though, should examine this direction. Finally, the proposed analysis raises interesting questions concerning the status of the implicated proposition *r* in *x but X*. The derivation of this implication seems similar to the way the derivation of 'real' scalar implicatures is usually described: The speaker uttered *everyone* and did not utter *(really) EVERYONE*, so we conclude that she uttered *everyone* to the refer to the exhaustified of the quantifier, i.e. to "Every in only the default / narrow domain, and not in the wide domain". However, since the conjuncts of ***but*** in this case are not scalar alternatives, i.e. neither of the alternatives is more informative / stronger than the other, the mechanism generating this implication does not seem to be covered by current approaches to scalar implicaure: Neither a 'pragmatic', neo Gricean mechanism, relying on the maxim of Quantity, nor a 'grammatical' mechanism (cf. Chierchia et al. (2011)), relying on the presence of an exhaustive operator in the syntax of the first conjunct. Future research, then, should examine the way such 'semi scalar implicatures' are generated.

References

Anscombre, J.-C. and Ducrot, O. C. (1984). L'argumentation dans la langue. philosophie et langage bruxelles: Pierre mardaga diteur, 1983. 184 p. *Dialogue*, 23(3):514517.

Charnavel, I. (2015). On scalar readings of french propre 'own'. *Natural Language & Linguistic Theory*, 34(3):807–863.

Chierchia, G. (2013). *Logic in Grammar*. Oxford University Press.

Chierchia, G., Danny, F., and Benjamin, S. (2011). The grammatical view of scalar implicatures and the relationship between semantics and pragmatics. In Portner, P., Maienborn, C., and von Heusinger, K., editors, *Semantics. An international handbook of contemporary research*. Mouton De Gruyter.

Davidson, D. (1979). Quotation. *Theory and Decision*, 11(1):27–40.

Fox, D. (2007). Free choice and the theory of scalar implicatures. In Sauerland, U. and Stateva, P., editors, *Presupposition and Implicature in Compositional Semantics*, pages 71–120. Palgrave Macmillan UK, London.

Fox, D. and Katzir, R. (2011). On the characterization of alternatives. *Natural Language Semantics*, 19(1):87–107.

Geurts, B. and Maier, E. (2003). Quotation in context. *Belgian Journal of Linguistics*, 17(1):109–128.

Ghomeshi, J., Jackendoff, R., Rosen, N., and Russell, K. (2004). Contrastive focus reduplication in English (the salad-salad paper). *Natural Language & Linguistic Theory*, 22(2):307–357.

Greenberg, Y. (2014). External and internal alternative-sensitive operators. An international workshop on Focus Sensitive Expressions from a Cross Linguistic Perspective, Bar Ilan University.

Greenberg, Y. (2015). Even, comparative likelihood and gradability. In Brochhagen, T., Roelofsen, F., and Theiler, N., editors, *Proceedings of the Amsterdam Colloquium 20*, pages 147–156. Amsterdam:UVA.

Greenberg, Y. (2016a). Metalinguistic contrast and scalar implicatures: The case of x but x. A paper presented at IATL32, Tel Aviv.

Greenberg, Y. (2016b). A new parameter for typologies of *even*-like operators: Operating over covert-based alternatives. A paper presented at SuB 21, Edinburgh.

Greenberg, Y. (2017). A revised, gradability-based semantics for *even*. *Natural Language Semantics*.

Greenberg, Y. and Orenstein, D. (2016). Typologies for *even*-like and *only*-like operators: Evidence from Modern Hebrew. A paper presented in ESSLLI 28, Bolzano.

Greenberg, Y. and Wolf, L. (2017). Gradable assertion speech acts. In *NELS48*. University of Reykyavick.

Hoeksema, J. (2007). Parasitic licensing of negative polarity items. *The Journal of Comparative Germanic Linguistics*, 10(3):163.

Jasinskaja, K. (2012). Correction by adversative and additive markers. *Lingua*, 122(15):1899 – 1918. SI: Additivity and Adversativity.

Jasinskaja, K. and Zeevat, H. (2008). Explaining additive, adversative and contrast marking in Russian and English. *Revue de Smantique et Pragmatique*, 24:65–91.

Kadmon, N. and Landman, F. (1993). Any. *Linguistics and Philosophy*, 16(4):353–422.

Katzir, R. (2014). On the roles of markedness and contradiction in the use of alternatives. In Reda, S. P., editor, *Pragmatics, Semantics and the Case of Scalar Implicatures*, pages 40–71. Palgrave Macmillan UK, London.

Kennedy, C. and McNally, L. (2005). Scale structure, degree modification, and the semantics of gradable predicates. *Language*, 81(2):345–381.

Maier, E. (2014). Mixed quotation: The grammar of apparently transparent opacity. *Semantics and Pragmatics*, 7(7):1–67.

Maier, E. (2017). Mixed quotations. to appear in. In Matthewson, L., Meier, C., Rullman, H., and Zimmermann, T. E., editors, *The Companion to Semantics*. Wiley Blackwell. [3rd version, Sept 2017].

Mayr, C. and Romoli, J. (2016). A puzzle for theories of redundancy: Exhaustification, incrementality, and the notion of local context. *Semantics and Pragmatics*, 9(7):1–48.

Merin, A. (1999). Information, relevance, and social decision-making: some principles and decision-theoretic semantics. In *Logic, Language and Computation: Volume 2*. Center for the Study of Language and Inf.

Partee, B. H. and Rooth, M. (2008). Generalized conjunction and type ambiguity. In *Formal Semantics*, pages 334–356. Blackwell Publishers Ltd.

Shan, C. (2010). The character of quotation. *Linguistics and Philosophy*, 33(5):417–443.

Shitrit, R. (2015). x aval X: A semantic-pragmatic analysis. Master's thesis, Bar-Ilan University.

Toosarvandani, M. (2014). Contrast and the structure of discourse. *Semantics and Pragmatics*, 7(4):1–57.

Umbach, C. (2005). Contrast and information structure: A focus-based analysis of but. *Linguistics*, 43(1).

von Fintel, K. (2004). How multi-dimensional is quotation? Handout, Harvard MIT UConn Workshop on Indexicals, Speech Acts & Logophors. http : //web.mit.edu/fintel/fintel-2004-pottsquotecomments.pdf.

Winter, Y. and Rimon, M. (1994). Contrast and implication in natural language. *Journal of Semantics*, 11(4):365–406.

Winterstein, G. (2013). The independence of quantity implicatures and contrast relations. *Lingua*, 132(Supplement C):67 – 84. SI: Implicature and Discourse Structure.

Wolf, L. (2015). Its probably certain. In *Proceedings of IATL30*. Ben-Gurion University of Negev.

Inverse Linking: Taking Scope with Dependent Types

Justyna Grudzińska[1] and Marek Zawadowski[2]

[1] Institute of Philosophy, University of Warsaw, Warsaw, Poland
j.grudzinska@uw.edu.pl
[2] Institute of Mathematics, University of Warsaw, Warsaw, Poland
zawado@mimuw.edu.pl

Abstract

Inverse linking constructions (ILCs) refer to complex DPs which contain a quantified NP (QP) which is selected by a preposition (e.g. *a representative of every country*). ILCs have been known to be ambiguous between a surface-scope reading and an inverse-scope reading. One puzzling difficulty for the existing accounts of ILCs is that some prepositions (e.g. *with*) block inverse-scope interpretations. In this paper, we propose a new dependent type analysis of the two readings of ILCs. In our analysis, we follow Zimmermann (2002) and assume that ILCs are structurally ambiguous at surface structure: the two readings of ILCs derive from the two (string identical) surface structures. The advantage of our dependent type account over Zimmermann's analysis is that it interprets the surface structure for the inverse-scope reading in a fully compositional way. Other compositional non-movement accounts of ILCs have been proposed; however, our dependent type account is the first to offer a principled solution to the puzzle of why inverse-scope readings are blocked with certain prepositions.

1 Introduction

ILCs refer to complex DPs which contain a quantified NP (QP) which is selected by a preposition, as illustrated in (1):

(1) A representative of every country is bald.

ILCs like (1) are considered to be ambiguous between a surface-scope reading and an inverse-scope reading.[1] On the surface-scope reading, (1) is understood to mean that there is some one person who represents every country and who is bald. On the inverse-scope reading, (1) is understood to mean that a different representative of each country is bald in each case. Puzzlingly, some prepositions (e.g. *with*) resist inverse-scope interpretations, as illustrated in (2):

(2) Someone with every known skeleton key opened this door.

Sentence (2) can only mean that there is some one person who happens to have every known skeleton key and who opened the door (for the discussion of the puzzle, see [27]). In this paper, we propose a new dependent type analysis of the two readings of ILCs.[2] Standard analyses of ILCs involve LF-movement [25, 26, 18, 16]. A more recent account is Zimmermann's surface-structure analysis [33]. In our analysis, we follow Zimmermann and assume that ILCs are

[1] In May's [25], ILCs were thought to allow inverse-scope readings only but this has changed since May's [26], and both surface-scope and inverse-scope readings are now considered to be available for ILCs (see e.g., [14, 15, 5, 32]).

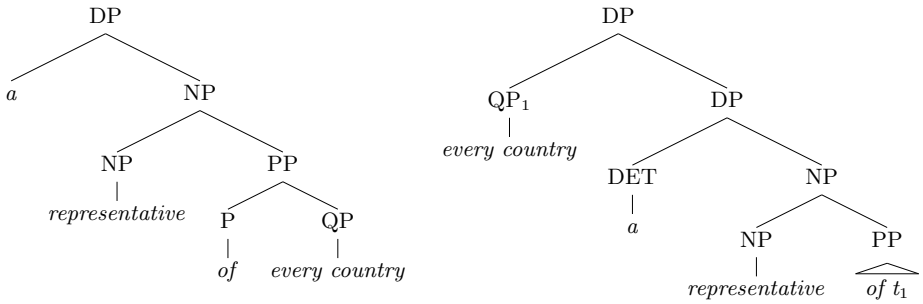
[2] Another characteristic property of ILCs is that, on the inverse-scope reading, the DP-internal QP (*every country*) can bind a pronoun outside the ILC, as in: A representative of every country despises it. The discussion of this property goes beyond the scope of this paper (but see [12] for our dependent type analysis of bound readings).

structurally ambiguous at surface structure: the two readings of ILCs derive from the two (string identical) surface structures. But whereas the surface structure for the inverse-scope reading has been interpreted using a non-fully compositional procedure in [33], we will show that it can be interpreted in a fully compositional way in our semantic framework with dependent types [10, 11]. Other compositional non-movement accounts of ILCs have been proposed: Hendriks's type-shifting approach [13] and Barker and Shan's continuation-based strategies [2, 3]. However, to the best of our knowledge, our dependent type account is the first to offer a principled solution to the puzzle of why inverse-scope readings are blocked with certain prepositions.

The structure of the paper is as follows. In section 2, we review Zimmermann's surface-structure analysis. Section 3 introduces the main features of our semantic framework with dependent types. In 4, we present a new dependent type analysis of the two reading of ILCs and show that the analysis proposed can account for the preposition puzzle, before concluding in 5.

2 ILCs and Surface-Structure Ambiguity

In [33], Zimmermann argues that the two readings of ILCs derive from the two different surface structures (SS's) and that there is syntactic evidence for distinguishing the two SS's. Standard LF-based approaches assume that the prepositional phrase (PP) in ILCs has the syntactic status of a regular postnominal modifier, i.e., the PP (*of every country*) stands in the sister position to the head noun (*representative*). The inverse-scope reading of ILCs is attributed to the application of quantifier raising (QR). QR replaces the QP *every country* with the coindexed trace (t_1), and adjoins it at DP [26, 18, 1]:[3]



The central observation that drives Zimmermann's surface-structure analysis is that PPs containing QPs that give rise to inverse-scope readings cannot freely change places with regular postnominal modifiers (e.g. relative clauses (RCs)) but must be DP-final, as illustrated by the examples below:

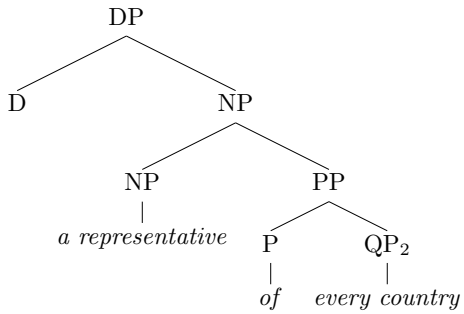
(3) One person [RC who was famous] [PP from every city] was invited.

(4) ♯ One person [PP from every city] [RC who was famous] was invited.

Sentence (3) can be understood to mean that every city x is such that one famous person from x was invited, while sentence (4) is semantically odd — it only allows a surface-scope reading saying that one person who came from every city and who was famous was invited. Inverse-scope readings are possible when PPs follow RCs (as in (3)), while non-final PPs give

[3] ILCs have been also analyzed as involving adjunction of the QP at S (see e.g., [25, 9]).

rise to surface-scope readings only (as in (4)). This asymmetry is unexpected on the LF-based analysis since all postnominal modifiers (PPs, RCs) have the same syntactic status. Based on this argument, Zimmermann proposes a different structure for the inverse-scope reading, where the PP (*of every country*) is not a regular postnominal modifier but is right-adjoined to the whole indefinite expression *a representative* (for this analysis, the adjectival theory of indefinite expressions is assumed, see e.g. [17]):



On the semantic side, the PP is reinterpreted as a generalized quantifier, as in the formula below:

$$\|of\ every\ country\| = \lambda P.\forall z[z \in country(z) \rightarrow \exists X[P(X) \wedge of(X, z)]],$$

where P stands for the property of pluralities denoted by the indefinite expression. Under this analysis, the observed problematic asymmetry can be readily explained. In the presence of regular postnominal modifiers (e.g. RCs), the PP that gives rise to the inverse-scope reading must be DP-final because its adjunction blocks the adjunction of additional regular postnominal modifiers. If the PP is followed by regular postnominal modifiers, it must be interpreted as a regular postnominal modifier itself and can give rise to the surface-scope reading only. The main problem for this account, as pointed out by Zimmermann himself, is that the reinterpretation of the PP is not a compositional semantic operation, i.e., it does not come as a result of the composition of the meanings of the constituent parts.

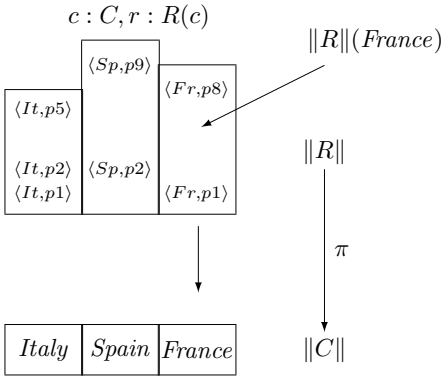
In our analysis, we adopt Zimmermann's position that ILCs are structurally ambiguous at surface structure (the two readings derive from the two surface structures distinguished by the syntactic position of the PP). The advantage of our dependent type account over Zimmermann's analysis is that it interprets the surface structure for the inverse-scope reading in a fully compositional way.

3 Semantics with Dependent Types

At the heart of our analysis is a dependent type theoretical framework [23, 24, 22, 6]. Whereas previous approaches adopted within the dependent type theoretical framework have been either proof-theoretic [31, 21, 4] or involved a combination of proof-theoretic and model-theoretic elements [8, 7], our semantics is model-theoretic with truth and reference being basic concepts (and no proofs). In this section, we introduce the main features of our semantic framework with dependent types [10, 11]. In 3.1, we introduce the concept of dependent types. In 3.2, it is shown how QPs and predicates are to be interpreted in this framework. In 3.3, we present a novel dependent type account of relational nouns.

3.1 Dependent Types

The two key features of our semantic framework are: many-sorted (many-typed) analysis and type dependency. Our analysis is many-sorted in the sense that it includes many basic types (e.g., type *M(an)*, type *C(ountry)*, ...). The variables of our semantic system are always typed. We write $c : C$ to denote that the variable c is of type C. Types are interpreted as sets. We write the interpretation of the type C as $\|C\|$. In our system, types can depend on the variables of other types, e.g., if c is a variable of the type of countries C, there is a type $R(c)$ of the representatives of that country



If we interpret type C as the set $\|C\|$ of countries, then we can interpret R as the set of the representatives of the countries in $\|C\|$, i.e. as the set of pairs:

$$\|R\| = \{\langle a, p\rangle : p \text{ is the person from the country } a\}$$

equipped with the projection $\pi : \|R\| \to \|C\|$. The particular sets $\|R\|(a)$ of the representatives of the country a can be recovered as the fibers of this projection (the preimages of $\{a\}$ under π):

$$\|R\|(a) = \{r \in \|R\| : \pi(r) = a\}.$$

Our semantic system makes no use of assignment functions; variables serve to determine dependencies and act as an auxiliary syntactic tool to determine how the operations combining interpretations of QPs and predicates are to be applied.

3.2 QPs and Predicates

QPs and predicates are interpreted relative to the context, where context is understood type-theoretically as a sequence of type specifications of the (individual) variables:

$$x : X, y : Y(x), z : Z(x, y), \ldots$$

Here, type Z depends on the variables x and y of types X and Y, respectively; type Y, on the variable x of type X; and type X is a constant type, i.e., it does not depend on any variables (for the formal definition of our type-theoretic notion of context, see [11]).

We assume a polymorphic interpretation of quantifiers. A generalized quantifier associates to every set Z a subset of the power set of Z:

$$\|Q\|(Z) \subseteq \mathcal{P}(Z).$$

Whereas in the standard Montagovian setting QPs are interpreted over the universe of all entities E [28], in our dependent type theoretical framework they are interpreted over types. For example, *some man* is interpreted over the type *Man* (given in the context), i.e., *some man* denotes the set of all non-empty subsets of the set of men:

$$\|\exists\|(\|Man\|) = \{X \subseteq \|Man\| : X \neq \emptyset\}.$$

In our semantic framework, quantification is also allowed over fibers, e.g., we can quantify existentially over the fiber of the representatives of France, as in *some representative of France*:

$$\|\exists\|(\|R\|(France)) = \{X \subseteq \|R\|(France) : X \neq \emptyset\}.$$

As a consequence of our many-sorted (many-typed) analysis, we also have a polymorphic interpretation of predicates. A predicate like *love* is interpreted over types (given in the context, e.g. *Man*, *Book*, ...), and not over the universe of all entities.

3.3 Relational Nouns

Nouns can be classified into three kinds: sortal (e.g., *man*, *country*, *book*), relational (e.g., *representative*, *part*, *attribute*) and functional (e.g., *mother*, *head*, *age*) (see. e.g., [19, 20]). In this paper, we only consider sortal and relational nouns. Whereas in the Montagovian setting sortal nouns are interpreted as one-place relations (expressions of type $\langle e, t\rangle$), in our dependent type theoretical framework they are treated as types. For example, *man* is interpreted as the type M/set of men. Nouns modified by regular postnominal modifiers are also interpreted as types. For example, *man who represents every city* is interpreted as the type/set of men who represent all the cities. In the Montagovian setting relational nouns (e.g. *representative*) are interpreted as two-place relations (expressions of type $\langle e, \langle e, t\rangle\rangle$). Our framework allows us to treat them as dependent types, e.g., *representative* (as in *a representative of a country*) is interpreted as the dependent type $c : C, r : R(c)$/for any element a in the set of countries $\|C\|$, there is a set (fiber) $\|R\|(a)$ of the representatives of that country. Our analysis can be further extended to nested relational nouns, e.g., *a formulation of a solution of a problem* can be interpreted as the nested dependency $p : P, s : S(p), f : F(p, s)$.

As discussed in Partee and Borschev's [29], the boundary between sortal and relational nouns is pervasively permeable. Sortal nouns can undergo 'sortal-to-relational' shifts, as in uses of sortal nouns with an overt argument (e.g., *book* expresses a sortal concept but its relational use can be coerced in *books of* ...). 'Relational-to-sortal' shifts are also possible, as in uses of relational nouns without an overt argument. Some nouns can even have both meanings, e.g. *child* can be understood to mean either 'direct descendant of' (relational reading) or 'nonadult' (sortal reading). For more discussion on sortal and relational nouns, see [19, 20, 29, 30]. The possibility of both 'sortal-to-relational' and 'relational-to-sortal' shifts has an important role to play in our account of ILCs.

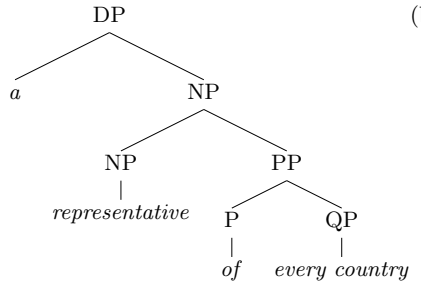
4 Dependent Type Analysis of ILCs

This section puts forth our new dependent type account of ILCs. In 4.1, we show how our account produces the two readings for ILCs. In 4.2, we provide the details of our compositional analysis of the inverse reading of ILCs. In 4.3, it is shown that the analysis proposed can account for the preposition puzzle.

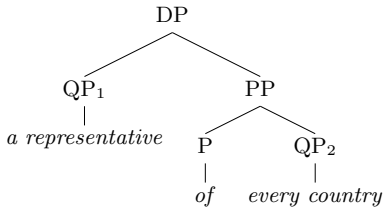
4.1 The Two Readings of ILCs

In our analysis, we follow Zimmermann in assuming that the two readings of ILCs derive from the two surface structures in (a) and (b) (we differ from Zimmermann in taking indefinites to be quantificational expressions but our proposal also can be made to work on the adjectival theory of indefinites):

(a)



(b)



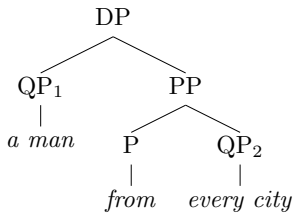
Under the perspective of our semantic framework with dependent types, the two structures in (a) and (b) (independently postulated and motivated by Zimmermann in [33]) come readily to mind — in (a) we quantify existentially over the single (complex) type/set of individuals who represent all the countries, while in (b) we quantify over two types, where the type *Representative* depends on (the variables of) the type *Country*.

More specifically, on the surface reading ((a)-structure), the PP stands in the sister position to the head nominal *representative*. Our proposal is that the relational noun *representative* undergoes a 'relational-to-sortal' shift when used with a regular postnominal modifier. The complex NP (noun modified by the postnominal PP) *representative of every country* is then interpreted as the type/set of individuals who represent all the countries and the DET *a* quantifies existentially over this set, yielding the surface ordering of quantifiers. On the inverse reading ((b)-structure), the PP is right-adjoined to the QP consisting of the head nominal (*a representative*). Crucially, the head nominal *representative* is now interpreted as the dependent type $c : C, r : R(c)$; the preposition *of* signals that *country* is a type on which *representative* depends; *country* is interpreted as the type C. By quantifying over $c : C, r : R(c)$, we get the inverse ordering of quantifiers (quantification over fibers is treated on a par with quantification over sets interpreting any other types):

$$\forall_{c:C}\exists_{r:R(c)}.$$

By making the type of representatives dependent on (the variables of) the type of countries our analysis forces the inversely linked reading without positing any extra scope mechanisms or arbitrary reinterpretation procedures.

One apparent problem for our proposal is that inverse-scope readings are also available for ILCs involving sortal nouns, as in *a man from every city*. Our solution to this problem is that 'sortal-to-relational' shifts are also possible and result in (b)-structures:



The relational use of sortal nouns (e.g. *man*) perhaps can be coerced by the implausibility of surface-scope readings with prepositions such as *from*, *in*, *on*. As discussed in [33], such prepositions specify the local position or origin of an entity and since entities do not occur at more than one place simultaneously, surface-scope readings of ILCs with these prepositions become implausible. Our analysis of the above structure is then exactly like the one just described for the inverse reading of *a representative of every country*.

4.2 The Compositional Analysis of the Inverse Reading of ILCs

This section explains the details of the compositional semantic analysis of the inverse-scope reading of ILCs on the example of sentence (1):

(1) A representative of every country is bald.

The complex DP *a representative of every country* is interpreted as the complex quantifier living on the set of all representatives. The interpretation of the structure:

$$c : C, r : R(c)$$

gives us access to the sets (fibers) $\|R\|(a)$ of the representatives of the particular country a only. To form the set of all representatives, we need to use type constructor Σ which takes the sum of fibers of representatives over countries in $\|C\|$. Thus the complex DP *a representative of every country* is interpreted as the complex quantifier living on $\|\Sigma_{c:C} R(c)\|$:

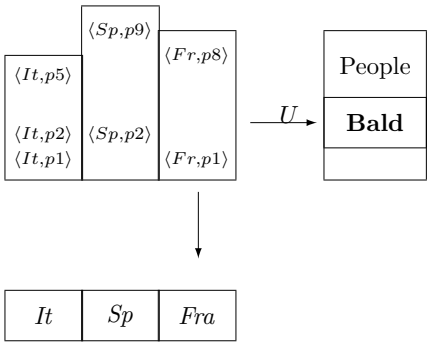
$$\|\forall_{c:C}\exists_{r:R(c)}\| = \{X \subseteq \|\Sigma_{c:C} R(c)\| : \{a \in \|C\| : \{b \in \|R\|(a) : b \in X\} \in \|\exists\|(\|R\|(a))\} \in \|\forall\|(\|C\|)\},$$

i.e., it denotes the set of the subsets of the set of representatives such that the set of countries such that each country has at least one representative in the corresponding fiber of representatives is the set of all countries.

Sentence (1) is true if and only if the set of bald representatives is in the denotation of the complex DP *a representative of every country*:

$$\|\forall_{c:C}\exists_{r:R(c)} Bald(r)\| = 1 \;\; \textit{iff} \;\; U^{-1}(\|Bald\|) \in \|\forall_{c:C}\exists_{r:R(c)}\|.$$

The illustration below serves to provide an intuitive understanding of the formula:



Note that a person counts as a representative only in virtue of standing in a particular relationship with some country. Now each representative has an underlying person, in fact two representatives can have the same underlying person (if this person represents two countries).

U stands for a function that forgets this part of the structure that relates people to countries and yields just the set of people. Predicate *Bald* is defined over the type/set of people — by taking the preimage of the set of bald people under function U, $U^{-1}(\|Bald\|)$, we get the set of bald representatives. For sentence (1) to be true, the set of bald representatives must be in the denotation of the complex quantifier expression *a representative of every country*.

4.3 Preposition Puzzle

ILCs are DPs which contain a QP which is selected by a preposition. In English this includes, among others, *of* (connector to a relational noun) and locative prepositions such as *at*, *from*, *in* or *on*. As discussed in May and Bale's [27], one puzzling difficulty for the existing accounts of ILCs is that some prepositions like *with* block inverse-scope readings. To illustrate the point, May and Bale give the example in (2):

(2) Someone with every known skeleton key opened this door.

Sentence (2) can only mean that one individual who happens to have every known skeleton key opened the door. Our solution to the preposition puzzle is that inverse-scope readings are unavailable for ILCs with prepositions which induce dependencies corresponding to the surface ordering of the QPs.

Note that in the case of prepositions like *of*, *from*, *in*, as in *a representative of (from, in) every country*, the 'dependent component' (*representative*) comes before the component on which it is dependent (*country*). Thus the dependency introduced $c : C, r : R(c)$ forces the inverse ordering of the QPs $\forall_{c:C}\exists_{r:R(c)}$. By contrast, in the case of prepositions like *with*, as in *a man with every key*, the 'dependent component' (*key*) comes after the component on which it is dependent (*man*). Thus the only possible dependency to be introduced is $m : M, k : K(m)$, one that corresponds to the surface ordering of the QPs $\exists_{m:M}\forall_{k:K(m)}$. Hence, under our analysis, the inverse-scope interpretation is unavailable to the QP in the object position of *with*.

As pointed out by an anonymous reviewer, examples like *a problem with every account* may present a difficulty for our proposal. Note, however, that *with* comes with a number of meanings, including: 'having or possessing (something)' and 'accompanied by; accompanying'. If the relation expressed is one of possession, as in our example (2), then the thing possessed depends on the possessor (as described above). If, however, the relation is that of accompanying, then the accompanying entity (problem) depends on the entity to be accompanied (account). Thus the dependency introduced is $a : A, p : P(a)$, forcing the inverse ordering of the QPs. Hence, under our analysis, the inverse-scope reading is available for the problematic example in question and, in general, for ILCs with the preposition *with* taken in the sense 'accompanying' (in line with intuitions reported by native speakers; consider also a slight variant of the reviewer's example found on the Internet: I had a problem with every game I played, from crashing to stupid errors).

5 Conclusion

In this paper, we provided a new dependent type analysis of ILCs. The advantage of our dependent type account over previous analyses (both LF-based approaches and Zimmermann's surface structure analysis) is that it is directly compositional. Using our semantic framework with dependent types, we can interpret the surface structure for the inverse-scope reading directly and in a fully compositional way. Furthermore, to the best of our knowledge, our dependent type account is the first to provide a principled solution to the preposition puzzle. Our proposal also

makes clear empirical predictions. Under Zimmermann's account, inverse-scope readings are expected to be a marked option (always a last resort), triggered by the implausibility of surface-scope readings with locative prepositions which leads to the reinterpretation procedure ([33]). Under our dependent type account, by contrast, inverse-scope readings should be a preferred option for relational nouns. In our future work, we plan to test empirically the implications of our proposal.

6 Acknowledgments

The research for this article is funded by the National Science Center on the basis of decision DEC-2016/23/B/HS1/00734. For valuable comments, the authors would like to thank the anonymous reviewers and the audience of *The Stockholm Logic Seminar* and two conferences: *The 18th Szklarska Poreba Workshop* and *The Workshop on Logic and Algorithms in Computational Linguistics 2017*.

References

[1] Chris Barker. Integrity: A syntactic constraint on quantificational scoping. In *Proceedings of the 20th West Coast Conference on Formal Linguistics*, pages 56–67, 2001.

[2] Chris Barker. Continuations and the nature of quantification. *Natural language semantics*, 10(3):211–242, 2002.

[3] Chris Barker and Chung-chieh Shan. *Continuations and natural language*, volume 53. Oxford Studies in Theoretical Linguistics, 2014.

[4] Daisuke Bekki. Representing anaphora with dependent types. In *International Conference on Logical Aspects of Computational Linguistics*, pages 14–29. Springer, 2014.

[5] Lucas Champollion and Uli Sauerland. Move and accommodate: A solution to haddock's puzzle. *Empirical Issues in Syntax and Semantics*, 8:27–51, 2011.

[6] Stergios Chatzikyriakidis and Zhaohui Luo. *Modern Perspectives in Type-Theoretical Semantics*. Springer, 2017.

[7] Robin Cooper. Dynamic generalised quantifiers and hypothetical contexts. *Ursus Philosophicus, a festschrift for Björn Haglund. Department of Philosophy, University of Gothenburg*, 2004.

[8] Tim Fernando. Conservative generalized quantifiers and presupposition. In *Proceedings of the 11th Semantics and Linguistic Theory Conference*, pages 172–191, 2001.

[9] Robert Fiengo and James Higginbotham. Opacity in np. *Linguistic Analysis*, 7(4):395–421, 1981.

[10] Justyna Grudzinska and Marek Zawadowski. System with generalized quantifiers on dependent types for anaphora. In *Proceedings of the EACL 2014 Workshop on Type Theory and Natural Language Semantics (TTNLS)*, pages 10–18, 2014.

[11] Justyna Grudzińska and Marek Zawadowski. Generalized quantifiers on dependent types: a system for anaphora. In *Modern Perspectives in Type-Theoretical Semantics*, pages 95–131. Springer, 2017.

[12] Justyna Grudzińska and Marek Zawadowski. Whence long-distance indefinite readings? solving chierchia's puzzle with dependent types. In *Proceedings of the 11th International Tbilisi Symposium on Logic, Language, and Computation, LNCS 10148*, pages 37–53. Springer, 2017.

[13] Herman Hendriks. *Studied flexibility: Categories and types in syntax and semantics*. Institute for Logic, Language and Computation, 1993.

[14] Yoon-kyoung Joh. Inverse-linking construction. *Language Research*, 44(2):275–297, 2008.

[15] Gregory M Kobele. Inverse linking via function composition. *Natural Language Semantics*, 18(2):183–196, 2010.

[16] Angelika Kratzer and Irene Heim. *Semantics in generative grammar*. Blackwell, 1998.

[17] Fred Landman. *Indefinites and the Type of Sets*. Wiley-Blackwell, 2008.

[18] Richard K Larson. Quantifying into np. *Ms. MIT*, 1985.

[19] Sebastian Löbner. Definites. *Journal of semantics*, 4(4):279–326, 1985.

[20] Sebastian Löbner. Concept types and determination. *Journal of semantics*, 28(3):279–333, 2011.

[21] Zhaohui Luo. Formal semantics in modern type theories with coercive subtyping. *Linguistics and Philosophy*, 35(6):491–513, 2012.

[22] Michael Makkai. First order logic with dependent sorts, with applications to category theory. *Preprint: http://www. math. mcgill. ca/makkai*, 1995.

[23] Per Martin-Löf. An intuitionistic theory of types: Predicative part. *Studies in Logic and the Foundations of Mathematics*, 80:73–118, 1975.

[24] Per Martin-Löf and Giovanni Sambin. *Intuitionistic type theory*, volume 9. Bibliopolis Napoli, 1984.

[25] Robert May. *The grammar of quantification*. PhD thesis, Massachusetts Institute of Technology, 1978.

[26] Robert May. *Logical Form: Its structure and derivation*, volume 12. MIT press, 1985.

[27] Robert May and Alan Bale. Inverse linking. In *The Blackwell companion to syntax*, pages 639–667. Blackwell Publishing, 2005.

[28] Richard Montague. The proper treatment of quantification in ordinary english. In *Approaches to natural language*, pages 221–242. Reidel Publishing Company, 1973.

[29] Barbara H Partee and Vladimir Borschev. Sortal, relational, and functional interpretations of nouns and russian container constructions. *Journal of Semantics*, 29(4):445–486, 2012.

[30] Stanley Peters and Dag Westerståhl. The semantics of possessives. *Language*, 89(4):713–759, 2013.

[31] Aarne Ranta. *Type-theoretical grammar*. Oxford University Press, 1994.

[32] Manfred Sailer. Inverse linking and telescoping as polyadic quantification. In *Proceedings of Sinn und Bedeutung 19*, pages 535–552, 2014.

[33] Malte Zimmermann. *Boys buying two sausages each: on the syntax and semantics of distance-distributivity*. PhD thesis, Netherlands Graduate School of Linguistics, 2002.

Causality and Evidentiality*

Yurie Hara

Waseda University and Hokkaido University
yuriehara@aoni.waseda.jp

Abstract

This paper formalizes the causal component of Davis & Hara's (2014) analysis of Japanese evidentiality, which defines "indirect evidence" as an observation of the *effect* state of the cause-effect dependency. The analysis correctly predicts that uttering *p-youda* only commits the speaker to 'if p, q must be true' but not to the prejacent p, and successfully derives the asymmetry between the prejacent p and the evidence source q.

1 Introduction

Japanese has an indirect evidential morpheme *youda*, which gives rise to (at least) two messages:

(1) Ame-ga futta youda.
rain-NOM fell EVID
'It seems that it rained.'
Message 1: "It rained." (M1)
Message 2: "The speaker has indirect evidence for 'it rained'." (M2)

Formal studies of evidentiality center around the following two questions: Q1. What are the statuses of the two messages? Q2. What is indirect evidence? Davis & Hara (2014) argued that unlike previous studies, M1 in (1) is an implicature while M2 is the assertional content of (1). Furthermore, Davis & Hara (2014) claim that indirect evidence for p is some state q which is usually caused by p. Thus, the at-issue content of (1) is that the speaker observed some state (say, wet streets) which is usually caused by a state which exemplifies the proposition "it rained". Davis & Hara's (2014) analysis overcomes the problems of the previous studies such as the lack of commitment to the prejacent and the evidential asymmetry, although the notion of causality is left as a primitive. The goal of this paper is to formally model the causality component in the interpretation of evidentials in the framework of causal premise semantics (Kaufmann, 2013).

This paper is structured as follows: We first review Davis & Hara's (2014) analysis which defines evidentiality as an observation of the effect state of the asymmetric causal relation in Section 2. To formalize the causality component in Davis & Hara's (2014) analysis, we review Kaufmann's (2013) causal premise semantics in Section 3.1. Section 3.2 demonstrates how the framework derives the evidential asymmetry. Section 4 concludes the paper.

2 Davis & Hara (2014)

2.1 p in *p-youda* is not an epistemic commitment

The previous studies on evidentials (Izvorski, 1997; Matthewson et al., 2006; McCready & Ogata, 2007; von Fintel & Gillies, 2010) predominantly argue that evidentiality is a kind of

*This work was supported by JSPS KAKENHI Grant Number 17H07172.

epistemic modality. That is, Evid(p) entails Modal(p). According to this line of analysis, since Modal(p) gives rise to an epistemic commitment to p, Evid(p) should also give rise to a commitment to p. In (2) and (3), to illustrate, both a bare assertion p and Modal(p) commit the speaker to p, thus p cannot be cancelled:

(2) #Ame-ga futta kedo jitsu-wa futtenai.
rain-NOM fell but fact-TOP fall-NEG
'#It rained but in fact it didn't.'

(3) #Ame-ga futta darou kedo jitsu-wa futtenai.
rain-NOM fell probably but fact-TOP fall-NEG
'#Probably, it rained but in fact it didn't.'

Thus, if an indirect evidential like *youda* were an epistemic modality, uttering *p-youda* should give rise to a commitment to p as well. However, Davis & Hara (2014) show that this treatment cannot be maintained since the prejacent p in *p-youda* is cancellable as in (4).[1]

(4) Ame-ga futta youda kedo, jitsu-wa futte-nai.
rain-NOM fell EVID but fact-TOP fall-NEG
'It seems that it rained, but in fact it didn't.'

In short, Davis & Hara (2014) conclude that the prejacent proposition is not an at-issue commitment of *p-youda* but a cancellable implicature.

2.2 What is indirect evidence?

McCready & Ogata (2007) treat evidentials as modals and offer a Bayesian semantics for evidentials, including *youda*. McCready & Ogata's account has the following two components:

(5) *McCready & Ogata's semantics of evidentials*:
p-youda, relativized to agent a, indicates that
1. some information q has led a to raise the subjective probability of p.
2. a takes p to be probably but not certainly true ($.5 < P_a(p) < 1$) after learning q.

According to McCready & Ogata (2007), thus, what counts as evidence is some information q that has led a to raise the subjective probability of p. In (6), a learns that the street is wet, which has led a to raise her subjective probability of p, hence the use of *youda* is acceptable.

(6) a. (Looking at wet streets)
b. Ame-ga futta youda.
rain-NOM fell EVID
'It seems that it rained.'

Although McCready & Ogata's theory provides a concrete way to define evidence and a reasonable analysis for (6), Davis & Hara (2014) show that it makes wrong predictions if we switch the prejacent proposition p and the evidence source q, as in (7). Learning that it is raining should also raise the agent's subjective probability of "the streets are wet", thus McCready & Ogata wrongly predict that *youda* is acceptable in (7).

[1] A similar argument is made for reportative evidentials by Faller (2002); Murray (2010); AnderBois (2014).

(7) a. (Looking at falling raindrops)
b. #Michi-ga nureteiru youda.
street-NOM wet EVID
'#It seems that the streets are wet.'

From this observation, Davis & Hara (2014) propose that indirect evidence for p is some event/state q that is usually caused by p. Wet streets are evidence for rain but rain is not evidence for wet streets because rain causes wet streets but not *vice versa*. Thus, in this analysis, *p-youda* is paraphrased as 'I perceived some underspecified event/state q which is caused by p.'

2.3 Evidentiality via Causality

From the observations discussed so far, Davis & Hara (2014) make the following two claims:

1. The prejacent p in *p-youda* is not an epistemic commitment but a cancellable implicature.
2. The semantics of *youda* needs to encode *asymmetric* causal dependencies, e.g., rain causes wet streets but not *vice versa*. In other words, what counts as evidence is an effect state of a cause-effect dependency.

Simply put, Davis & Hara (2014) define the interpretation of *p-youda* as follows:

(8) *Davis & Hara's interpretation of evidentials*
Evid(p) is true at w iff $\exists q$ such that the speaker perceives a state q at w and p causes q.

Under this analysis, (6) is paraphrased as 'I perceived a wet street and rain causes wet streets.' while (7) is paraphrased as 'I perceived rain and wet streets cause rain.' Since we know that the latter proposition is false according to our world knowledge, (7) is infelicitous.[2]

Although Davis & Hara (2014) overcome the problems of the previous analyses, the notion of causality is left unanalyzed as can be seen in (8). This paper formally implements the notion of causality in Kaufmann's (2013) causal premise semantics.

3 Formalizing the Causal Asymmetry

My analysis of interpretation of evidentials is built on the interpretation of conditional or modal statements. The important difference from the previous evidential-as-modal analyses is that the prejacent proposition p of Evid(p) contributes to the conditional antecedent or the modal restriction but not to the conditional consequent or the nuclear scope of the modal quantification. In a nutshell, I propose the following definition: Let M_c, O_c, w be a causal modal base, a causal ordering source O_c,and a possible world, respectively. An evidential statement Evid(p) is true at M_c, O_c, w just in case the speaker perceives some event/state q and if p is true, then q must be true at M_c, O_c, w:

(9) Proposal: *Interpretation of evidentials* (first version)
Evid(p) is true at M_c, O_c, w iff $\exists q$ such that the speaker perceives q at w and $\text{Must}_p(q)$ is true at M_c, O_c, w.

[2] See Sawada (2006); Takubo (2007) for similar analyses. Sawada (2006) argues that *youda* is a modality that infers a cause. Takubo (2007) claim that *youda* is attached to a proposition that is abductively inferred. See Davis & Hara (2014) for an argument against Takubo's analysis.

The following subsections present the technical preliminaries that are necessary to implement the proposal. In particular, the modal base and the ordering source need to be causally structured in order to capture the evidential asymmetry discussed above.

3.1 Technical Preliminaries

Kaufmann (2013) introduces causal networks to Kratzer's (2005, among others) premise semantics to interpret counterfactuals. I claim that the same apparatus can predict interpretations of evidentials.

3.1.1 Premise Sets and Structures

Kaufmann's framework extends Kratzerian premise semantics by deriving and ranking premise sets. A brief review of Kratzer's premise semantics is in order. Conversational backgrounds, f, g, are functions that map possible worlds to sets of propositions. Following the linguistic convetion, I use $f(w)$ and $g(w)$ for the modal base and the ordering source, respectively.

Modal statements are interpreted relative to premise sets. The premise sets are consistent set of propositions obtained by adding propositions from the ordering source to the modal base:

(10) *Kratzer premise sets*
Let M, O be two sets of propositions. The set $Prem^K(M, O)$ of Kratzer premise sets contains all and only the consistent supersets of M obtained by adding (zero or more) propositions from O. (Kaufmann, 2013, 1141)

If p is entailed by every possible premise set constructed from M and O, p is a necessity relative to M and O. In contrast, if p is consistent with some of the premise sets, p is a possibility relative to M and O:

(11) *Kratzer necessity and possibility*
Let Φ be a set of premise sets and p a proposition.
a. p is a necessity relative to Φ iff every premise set in Φ has a superset in Φ of which p is a consequence.
b. p is a possibility relative to Φ iff there is some premise set in Φ such that p is consistent with all of its supersets in Φ . (Kaufmann, 2013, 1141)

Now, modal statements can be interpreted as follows:

(12) *Kratzer interpretation of modals*
a. Must(p) is true at f, g, w iff p is a necessity relative to $\text{Prem}^K(f(w), g(w))$.
b. May(p) is true at f, g, w iff p is a possibility relative to $\text{Prem}^K(f(w), g(w))$.
(Kaufmann, 2013, 1141)

To illustrate, let us have a modal base $f(w) = \{p, q\}$ and an ordering source $g(w) = \{r, \bar{d}\}$. Then, we obtain the set of premise sets as in (13).

(13) $\text{Prem}^K(f(w), g(w)) = \{\{p, q\}, \{p, q, r\}, \{p, q, \bar{d}\}, \{p, q, r, \bar{d}\}\}$

Now, in order to capture the causal asymmetry, we introduce structures to the premise sets. Let f be a Kratzerian conversational background, which is a function from possible worlds to sets of propositions. Then, **f** is a *premise background* which is a function from possible worlds

to sets of sets of propositions defined in (14). Note that $\mathbf{f}$ alone contains no more or less information than f.

(14) A *premise background* $\mathbf{f}$ *structures* a Kratzerian conversational background f iff at all worlds v, $\mathbf{f}(v)$ is a set of subsets of $f(v)$. (Kaufmann, 2013, 1144)

To introduce the ranking of premise sets, a set of *sequence structures* is recursively defined as in (15). "$\leq_1 \times \leq_2$"signifies the lexicographic order on the Cartesian product.

(15) a. If Φ is a set of sets of propositions, then $\langle \Phi, \leq \rangle$ is a (basic) sequence structure.
b. If $\langle \Phi_1, \leq_1 \rangle$ and $\langle \Phi_2, \leq_2 \rangle$ are sequence structures, then so is $\langle \Phi_1, \leq_1 \rangle * \langle \Phi_2, \leq_2 \rangle$, defined as $\langle \Phi_1 \times \Phi_2, \leq_1 \times \leq_2 \rangle$. (Kaufmann, 2013, 1146)

A premise structure that is used for interpreting modal sentences is the set of *consistent* sequence structures (16).

(16) *Premise structure*:
$\mathrm{Prem}(\langle \Phi, \leq \rangle)$ is the pair $\langle \Phi', \leq' \rangle$, where Φ' is the set of consistent sequences in Φ and $\leq'$ is the restriction of $\leq$ to Φ'. (Kaufmann, 2013, 1147)

Modal statements can be interpreted in a way parallel to the Kratzer interpretation (12):

(17) *Interpretation of modals*:
a. $\mathrm{Must}(q)$ is true at $\mathbf{f}, \mathbf{g}, w$ iff q is a necessity relative to $\mathrm{Prem}((\mathbf{f}^* \mathbf{g})(w))$.
b. $\mathrm{May}(q)$ is true at $\mathbf{f}, \mathbf{g}, w$ iff q is a possibility relative to $\mathrm{Prem}((\mathbf{f}^* \mathbf{g})(w))$.
(Kaufmann, 2013, 1148)

To illustrate, given our conversational backgrounds, $\mathbf{f}(w) = \{\{p, q\}\}$ and $\mathbf{g}(w) = \{\emptyset, \{r\}, \{\bar{d}\}, \{r, \bar{d}\}\}$, we have the following set of premise structures (I follow Kaufmann's notation for readability: "$xy.$" stands for "$(\{x, y\}, \emptyset)$".):

(18) $\mathrm{Prem}(\mathbf{f} * \mathbf{g}(w)) = \mathbf{f}(w) \times \mathbf{g}(w) = \{(\{p, q\}, \emptyset), (\{p, q\}, \{r\}), (\{p, q\}, \{\bar{d}\}), (\{p, q\}, \{r, \bar{d}\})\}$
$= \{pq., pq.r, pq.\bar{d}, pq.r\bar{d}\}$

3.1.2 Causal Premise Semantics

Now we are ready to introduce causal structures (19) to capture causal asymmetries. A causal network has two components. The first part is a *directed acyclic graph* (DAG) in which vertices represent variables/partitions (e.g., R and D in Figure 1 representing "whether it is raining" and "whether streets are dry", respectively) and edges represent causal influence. The second component is that only the values of its immediate parents influence each variable.

Figure 1: Rain and Dry street

(19) A *causal structure* for non-empty W is a pair $\mathcal{C} = \langle U, < \rangle$, where U is a set of finite partitions on W and $<$ is a directed acyclic graph over U. (Kaufmann, 2013, 1151)

As with Kaufmann, I assume that causal dependency is not deterministic: the values of its parents determine whether the value of each variable is a necessity or a possibility (17). Furthermore, the premise background constrained by the ordering source determines the value of parents.

In order to interpret evidentials, I follow Kaufmann (2013) and postulate a causal premise background $\mathbf{f}_c$. The causal premise background $\mathbf{f}_c$ consists of causally relevant truths (20-b).

(20) a. The set Π^U of *causally relevant propositions* is the set of all cells of all partitions in U.
b. The set of *causally relevant truths* at w: $\Pi^U_w = \{p \in \Pi^U | p \text{ is true at } w\}$
(^{U} is omitted hereafter.) (Kaufmann, 2013, 1152)

Furthermore, $\mathbf{f}_c$ is constrained by the closure under ancestors, which ensures the *asymmetric* relation between variables X and Y. We need to introduce two notions to define the closure under ancestors, *setting* and *descendant*:

(21) a. A variable X is *set* in a set of propositions P iff exactly one of X's cells is in P.
b. X is a *descendant* of Y iff there is a path from Y to X of zero or more steps along the direction of causal influence. (Kaufmann, 2013, 1153)

Closure under ancestors is defined as follows:

(22) A subset P' of P is *closed under ancestors* in P iff [for all $X, Y \in U$ such that X is a descendant of Y and both are set in P], [if X is set in P', then Y is also set in P']. (Kaufmann, 2013, 1153)

To illustrate, let us consider $P = \{r, \bar{d}\}$ with the causal network depicted in Figure 1. Subsets $\emptyset$ and $\{r\}$ are closed under ancestors because D is not set in $\emptyset$ and $\{r\}$. $\{\bar{d}\}$ is not closed under ancestors because D is set but R is not set. $\{r, \bar{d}\}$ is closed under ancestors because D is set and so is R.

Taken together, $\mathbf{f}_c$ is postulated as in (23).

(23) $\mathbf{f}_c(w) := \{X \subseteq \Pi_w | X \text{ is closed under ancestors in } \Pi_w\}$ (Kaufmann, 2013, 1153)

Thus, if we have causal relevant truths $\Pi_w = \{r, \bar{d}\}$, our causal premise background is: $\mathbf{f}_c(w) = \{\emptyset, \{r\}, \{r, \bar{d}\}\}$.

Also, the other premise background, i.e., the ordering source $\mathbf{g}$ satisfies the *Causal Markov condition* relative to a causal structure $\mathcal{C}$. To define Causal Markov condition, a brief introduction to Conditional Independence is in order. The idea is the following: Consider a partition $X \in U$ and sets of partitions $\mathbf{Y}, \mathbf{Z} \subseteq U$. X is conditionally independent of $\mathbf{Y}$ given $\mathbf{Z}$ under $\mathbf{g}(w)$ if and only if learning the setting of $\mathbf{Y}$ in $\mathbf{Z}$ does not alter the value of any cells in X:

(24) *Conditional independence*
Let $\mathbf{g}$ be a premise background, w a possible world, and U a set of partitions. For any $X \in U$ and disjoint sets $\mathbf{Y}, \mathbf{Z} \subseteq U$ not containing X: X is *conditionally independent* of $\mathbf{Y}$ given $\mathbf{Z}$ under $\mathbf{g}(w)$ iff for all cells $x \in X$, partial settings $\mathbf{y}$ of $\mathbf{Y}$ and settings $\mathbf{z}$ of $\mathbf{Z}$ such that $\mathbf{y} \cup \mathbf{z}$ is consistent, x is a necessity (possibility) relative to $\text{Prem}(\{\mathbf{z}\} * \mathbf{g}(w))$ iff x is a necessity (possibility) relative to $\text{Prem}(\{\mathbf{z} \cup \mathbf{y}\} * \mathbf{g}(w))$
(Kaufmann, 2013, 1155)

The idea behind the Causal Markov condition is that any partition X in the causal structure is independent of any of X's ancestors except for X's immediate parents. Let $pa(X)$ be the set

of X's parents and $de(X)$ be the set of X's descendants. Causal Markov condition is defined as follows:

(25) *Causal Markov condition*
Let $\mathcal{C} = \langle U, < \rangle$ be a causal structure and $\mathbf{g}$ a premise background. $\mathbf{g}$ satisfies the **Markov condition** relative to C if and only if for all $w \in W$ and $X \in U$, X is conditionally independent of $U \backslash (de(X) \cup pa(X))$, given $pa(X)$, under $\mathbf{g}(w)$.
(Kaufmann, 2013, 1156)

3.2 Deriving evidentiality from causality

Our interpretation of evidentials is built on the general interpretation of conditionals. In the current framework, we obtain a premise background $\mathbf{f}[p]$ by hypothetically updating a premise background $\mathbf{f}$ with the antecedent proposition p:

(26) *Hypothetical update*
For all w: $\mathbf{f}[p](w) := \{\{p\}\}^{*}\mathbf{f}(w)$. (Kaufmann, 2013, 1148)

For example, consider our pre-update modal base $\mathbf{f}(w)$ as in (27-a). We acquire the post-update modal base $\mathbf{f}[r](w)$ by appending the hypothetical proposition r before each member of $\mathbf{f}(w)$:

(27) a. $\mathbf{f}(w) = \{\emptyset, \{p\}, \{p, q\}\}$
b. $\mathbf{f}[r](w) = \{\{r\}\} * \{\emptyset, \{p\}, \{p, q\}\} = \{r., r.p, r.pq\}$

Finally, we define the interpretation of evidentials. Evid(p) is true at $\mathbf{f}_c, \mathbf{g}, w$ when there is some state q such that the speaker perceives q at w and q is a necessity relative to $\text{Prem}((\mathbf{f}_c[p]^{*}\mathbf{g})(w))$:

(28) *Interpretation of evidentials* (final version)
Evid(p) is true at $\mathbf{f}_c, \mathbf{g}, w$ iff $\exists q$ such that the speaker perceives q at w and $\text{Must}_p(q)$ is true at $\mathbf{f}_c, \mathbf{g}, w$.

Note that the prejacent p of *p-youda* contributes to the antecedent rather than the consequent. In other words, Evid(p) only commits the speaker to $\text{Must}_p(q)$ and not to Must(p).

Let us illustrate the working of (28). Consider now the three-variable network in Figure 2 (the variable H represents "whether water is hose-sprayed") with the causally relevant propositions $\Pi = \{r, \bar{r}, h, \bar{h}, d, \bar{d}\}$.

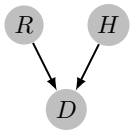


Figure 2: Rain, water-Hose and Dry street

Let us take example (4) repeated here as (29) first. (29) shows that the speaker can felicitously deny the prejacent proposition of the evidential statement.

(29) Ame-ga futta youda kedo, jitsu-wa futte-nai.
rain-NOM fell EVID but fact-TOP fall-NEG
'It seems that it rained, but in fact it didn't.'

Suppose that at w, it is not raining ($\bar{r}$), water is hose-sprayed (h), and streets are wet ($\bar{d}$) as in (30-a). By (23), we obtain (30-b). Note that $\{\bar{d}\} \notin \mathbf{f}_c(w)$, that is, $\{\bar{d}\}$ is not closed under ancestors in Π_w since D is set but R and H are not set in $\{\bar{d}\}$. Next, we obtain the premise structure (30-c) by hypothetically updating $\mathbf{f}_c(w)$ with r and removing any inconsistent sequences. As for the ordring source, let us assume that $\mathbf{g}$ prescribes that normally, water is not hose-sprayed, rain implies wet streets and so does hose-spraying (30-d). From (30-c) and (30-d), we obtain (30-e). Since $\bar{d}$ is a necessity relative to $\mathrm{Prem}((\mathbf{f}_c[r]^*\mathbf{g})(w))$, $\mathrm{Must}_r(\bar{d})$ is true at $\mathbf{f}, \mathbf{g}, w$.

(30) a. $\Pi_w = \{\bar{r}, h, \bar{d}\}$
b. $\mathbf{f}_c(w) = \{\emptyset, \{\bar{r}\}, \{h\}, \{\bar{r}, h\}, \{\bar{r}, h, \bar{d}\}\}$
c. $\mathrm{Prem}(\mathbf{f}_c[r](w)) = \{r., r.h\}$
d. $\mathbf{g}(w) = \{\emptyset, \{\bar{h}\}, \{r \to \bar{d}\}, \{h \to \bar{d}\}\}$
e. $\max \mathrm{Prem}((\mathbf{f}_c[r]^*\mathbf{g})(w)) = \{r.h.(r \to \bar{d}), r.h.(h \to \bar{d})\}$

As a result, given that the speaker perceives $\bar{d}$ at w, $\mathrm{Evid}(r)$ is true at $\mathbf{f}_c, \mathbf{g}, w$ because $\mathrm{Must}_r(\bar{d})$ is true at $\mathbf{f}_c, \mathbf{g}, w$, even though r is not true at w, Thus, (29) can be uttered felicitously without the speaker's commitment to r.

Finally, let us see if (28) can explain the evidential asymmetry. Let us derive the felicitous interpretation of (6) repeated here as (31) first.

(31) a. (Looking at wet streets)
b. Ame-ga futta youda.
rain-NOM fell EVID
'It seems that it rained.'

Suppose that the causally relevant truths at v are as in (32-a). Note that $\mathbf{g}(w) = \mathbf{g}(v)$.

(32) a. $\Pi_v = \{r, \bar{h}, \bar{d}\}$
b. $\mathbf{f}_c(v) = \{\emptyset, \{r\}, \{\bar{h}\}, \{r, \bar{h}\}, \{r, \bar{h}, \bar{d}\}\}$
c. $\mathrm{Prem}(\mathbf{f}_c[r](v)) = \{r., r.r, r.\bar{h}, r.r\bar{h}, r.r\bar{h}\bar{d}\}$
d. $\mathbf{g}(v) = \{\emptyset, \{\bar{h}\}, \{r \to \bar{d}\}, \{h \to \bar{d}\}\}$
e. $\max \mathrm{Prem}((\mathbf{f}_c[r]^*\mathbf{g})(v)) = \{r.\bar{h}.(r \to \bar{d}), r.\bar{h}.(h \to \bar{d}), r.r\bar{h}\bar{d}.(r \to \bar{d}), r.r\bar{h}\bar{d}.(h \to \bar{d})\}$

Assuming that the speaker perceives $\bar{d}$ at v, $\mathrm{Evid}(r)$ is true at $\mathbf{f}_c, \mathbf{g}, v$ because $\bar{d}$ is a necessity relative to $\mathrm{Prem}((\mathbf{f}_c[r]^*\mathbf{g})(v))$, hence $\mathrm{Must}_r(\bar{d})$ is true at $\mathbf{f}_c, \mathbf{g}, v$. The speaker of (31) is asserting that she perceived wet streets and if it rains, streets must be wet, i.e., rain causes wet streets.

Finally, we derive the infelicity of (7), repeated here as (33).

(33) a. (Looking at falling raindrops)
b. #Michi-ga nureteiru youda.
street-NOM wet EVID
'#It seems that the streets are wet.'

Take the same world v as above, thus we have the same causally relevant truths Π_v and causal modal base $\mathbf{f}_c(v)$ as the ones in (32). Now (33) translates to $\mathrm{Evid}(\bar{d})$, so the modal base is altered by hypothetically updating $\mathbf{f}_c(v)$ with $\bar{d}$ as in (34-c). Together with the ordering

source $\mathbf{g}(v)$, Evid($\bar{d}$) is interpreted relative to the set of premise structures shown in (34-e).

(34) .

a. $\Pi_v = \{r, \bar{h}, \bar{d}\}$
b. $\mathbf{f}_c(v) = \{\emptyset, \{r\}, \{\bar{h}\}, \{r, \bar{h}\}, \{r, \bar{h}, \bar{d}\}\}$
c. $\text{Prem}(\mathbf{f}_c[\bar{d}](v)) = \{\bar{d}., \bar{d}.r, \bar{d}.\bar{h}, \bar{d}.r\bar{h}\}$
d. $\mathbf{g}(v) = \{\emptyset, \{\bar{h}\}, \{r \to \bar{d}\}, \{h \to \bar{d}\}\}$
e. $\max \text{Prem}((\mathbf{f}_c[\bar{d}]^*\mathbf{g})(v)) \supseteq \{\bar{d}.\bar{h}.(r \to \bar{d}), \bar{d}.\bar{h}.(h \to \bar{d})\}$

Evid($\bar{d}$) is *not* true at $\mathbf{f}_c, \mathbf{g}, w$ because r is is not a necessity relative to $\text{Prem}((\mathbf{f}_c[\bar{d}] * \mathbf{g})(v))$, i.e., $\text{Must}_{\bar{d}}(r)$ is false at $\mathbf{f}_c, \mathbf{g}, v$. Put another way, (33) is infelicitous since the speaker is making a false claim, 'wet streets cause rain' (even if the speaker does observe rain).

3.3 Summary

We offer a formal implementation of the causal component of evidentiality which correctly predicts the lack of commitment to the prejacent proposition and derive the asymmetry of the evidential dependency between the prejacent and the evidence source from the asymmetry between the ancestor and the descendent in a causal network.

4 Conclusion

We formalized the causal component of Davis & Hara's (2014) analysis of Japanese evidentiality, which defines "indirect evidence" as the *effect* state of the cause-effect dependency, correctly predicts that uttering *p-youda* only commits the speaker to $\text{Must}_p(q)$ but not to the prejacent p, and successfully derive the asymmetry between the prejacent p and the evidence source q.

References

AnderBois, Scott. 2014. On the exceptional status of reportative evidentials. In *Proceedings of SALT 24*, 234–254.

Davis, Christopher & Yurie Hara. 2014. Evidentiality as a causal relation: A case study from Japanese *youda*. In Christopher Pi nón (ed.), *Empirical Issues in Syntax and Semantics 10*, .

Faller, Martina. 2002. *Semantics and Pragmatics of Evidentials in Cuzco Quechua*: Stanford University dissertation.

von Fintel, Kai & Anthony S. Gillies. 2010. Must . . . stay . . . strong! *Natural Language Semantics* 18(4). 351–383. doi:10.1007/s11050-010-9058-2.

Izvorski, Roumyana. 1997. The Present Perfect as an Epistemic Modal. *the proceedings of SALT* 7. 222–239.

Kaufmann, Stefan. 2013. Causal premise semantics. *Cognitive Science* 37. 1136–1170.

Kratzer, A. 2005. Constraining premise sets for counterfactuals. *Journal of Semantics* 22. 153–158.

Matthewson, Lisa, Hotze Rullmann & Henry Davis. 2006. Evidentials are epistemic modals in St'át'imcets. In Masaru Kiyota, James L. Thompson & Noriko Yamane-Tanaka (eds.), *Papers for the 41st International Conference on Salish and Neighbouring Languages* University of British Columbia Working Papers in Linguistics 18, 221–263.

McCready, Eric & Norry Ogata. 2007. Evidentiality, modality and probability. *Linguistics and Philosophy* 30(2). 35–63.

Murray, Sarah E. 2010. *Evidentiality and the Structure of Speech Acts*: Rutgers dissertation. http://www.semanticsarchive.net/Archive/WViOGQxY/.

Sawada, Harumi. 2006. *Modaritii.* Kaitakusha.

Takubo, Yukinori. 2007. Two types of modal auxiliaries in japanese: Two directionalities in inference. In *Japanese/Korean Linguistics*, vol. 15, University of Chicago Press.

May or *Might*? Strength, Duality and Social Meaning*

Hadil Karawani[1] and Brandon Waldon[2]

[1] Leibniz-ZAS, Berlin, Germany karawani@leibniz-zas.de
[2] Stanford University, Stanford, California, U.S.A. bwaldon@stanford.edu

This paper addresses one longstanding claim about epistemic *must*, namely that its semantic dual is the modal auxiliary *might*. While this is a "usual assumption" in the natural language semantics literature (Lassiter 2016: 14), it is not universally accepted: Crespo, Karawani, and Veltman (2017) (henceforth CKV) propose a theory of epistemic modality whereby the dual of *must* is *may*, the non-subjunctive counterpart of *might*. According to their view, an assymetric entailment relation holds between *may* and *might*, whereby *might* is weaker than *may* with respect to speakers' expectations about the likelihood of the prejacent. Consistent with this view, the results of one experiment suggest that English speakers consider *may p* to be stronger than *might p* on two metrics: perceived speaker certainty of *p* and inferred likelihood of *p*. In two subsequent experiments, we were unable to find conclusive evidence that speakers distinguish between epistemic *may* and *might* in discourse contexts, including in epistemic contradiction contexts of the form *must p, but might/may not p*. We interpret this as evidence that the "usual assumption" at a minimum needs to be revised to accommodate *may*. We conclude with a discussion of the diachronic, cross-linguistic, and social meaning facets of epistemic modality that are uniquely accounted for by the framework of CKV.

1 Background

Much of the recent literature on epistemic modality in English has centered around the strength of *must p*, which on some accounts is claimed to be true iff *p* is true in all epistemically possible worlds (c.f. von Fintel & Gillies 2010, Willer 2013); and the strength of *might p*, which on some accounts is claimed to be true so long as there is one epistemically accessible world in which *p* is true (c.f. Rudin 2016). These two claims are theoretically attractive: if the first is correct, then *must* is the natural language expression corresponding to the epistemic necessity operator of modal logic; if the second is correct, then *might* is the natural language counterpart of epistemic possibility. If we accept these claims in tandem, then the infelicity of the assertions in (1) can easily be explained in terms of semantic ill-formedness. Namely, (1) is infelicitous for precisely the same reasons that (2) is a contradiction in modal logic:

(1) #It must be raining, but it might not be raining.

(2) $\Box$ A $\wedge$ $\Diamond$ $\neg A$

Namely, *must* and *might*, along with $\Box$ and $\Diamond$, are semantic duals: that is, *might p* is semantically equivalent to $\neg$*must*$\neg p$, just as $\Diamond$A is semantically equivalent to $\neg\,\Box\,\neg$A. However,

*We thank Heather Burnett, Cleo Condoravdi, Judith Degen, Patrick Elliot, Anastasia Giannakidou, Nicole Gotzner, Stefanie Jannedy, Manfred Krifka, Beth Levin, Guillermo del Pinal, Josep Quer, Uli Sauerland, Stephanie Solt, Jack Tomlinson, Hubert Truckenbrot, Frank Veltman, and Hedde Zeijlstra, as well as audiences at the semantics reading group at ZAS and at the Words With Friends reading group at Stanford for generous and helpful feedback. All remaining mistakes are, of course, our own. We also thank Alexandre Cremers, Floris Roelofsen and the anonymous reviewers of Amsterdam Colloquium.

this is not the only analysis that provides us a story of the infelicity of (1) via semantic dualism: a variety of accounts on the market achieve this by strengthening the semantics of *might* and correspondingly weakening the semantics of *must*: for example, Kratzer's (1991) denotation of *must* is universal quantification over only the maximally-normal subset of epistemically accessible worlds, while *might* is existential quantification over this same set; Lassiter (2016) gives *must* and *might* a probabilistic semantics, where the probability threshold for *must* is defined as the inverse of that of *might*. The intuition, nevertheless, is still that the semantic duality relationship of *must* and *might* explains the contradictory feel of the sentences in (1). Willer (2013), on the other hand, argues for a strong denotation of *must* according to which *must p* entails *p*, but he also argues that *might* is strong: *might p* expresses that *p* is a live possibility, with 'live possibilities' defined as "possibilities that are compatible with the agent's evidence and that the agent takes seriously" (2013: 5). In his dynamic semantics of *must* and *might*, they are formally duals, and (1) is predicted to be infelicitous: if a context update of *must p* is incompatible with a subsequent update that *p* is epistemically possible (given that *p* is entailed), then *must p* is certainly incompatible with the suggestion that *p* is a possibility that we should seriously consider in future deliberations. Conversely, Veltman (1986) argues for a weak *must* (in which *must p* does not entail *p*) but also, in Veltman (1996), for a weak *might* (in which *might* is simply existential quantification over epistemically possible worlds). Note that Veltman does not treat *must* and *might* as duals. Moreover, (1) on Veltman's account is not predicted to express a semantic contradiction; the infelicity must be explained through other extra-semantic means. What is the dual of *must*, though, if not *might*? On Willer's (2013) story, another conceivable dual is *possible*, assuming that *possible* means something along the lines of it being only a bare epistemic possibility that the prejacent is true. This solution seems less tenable for weak theories of *must*: if *must* explicitly leaves open the epistemic possibility of the negation of the prejacent, then it would be surprising to find that *possible* is its dual (again assuming this rough sketch of the semantics of *possible*). In fact, in arguing for weak *must*, Lassiter (2016) provides numerous examples from the web where speakers make statements to the effect of *must p, but possible not p*. Given a weak theory of *must*, the dual (if it exists) needs to be stronger than bare possibility and thus intuitively will be a lexical item whose semantics are stronger than those of *possible*.

2 Introducing *may*

CKV propose a dynamic analysis of epistemic modality that builds on the assumption that *might* is weaker than *may*: in their theory, *might* expresses existential quantification over epistemically possible worlds, while *may* expresses existential quantification only over the *most likely* of those epistemically possible worlds – those that are consistent with one's expectations.[1] Hence, two kinds of possibility are introduced, with a distinction between knowledge and expectation:
May: *likely possibility, something to reckon with, consistent with one's expectations.*
Might: *unlikely possibility, consistent with everything one knows but not with expectations.*

In CKV's framework, *may*, and not *might*, is the dual of *must*, as *must* according to their analysis expresses universal quantification over the set over which *may* expresses existential quantification. One empirical prediction of this view is that while (1) is not necessarily contradictory, (3) most certainly is, for the familiar reason that the sentence contains both a lexical

[1] The relative strength of *may* and *might* is derived in the following way: *might* is equal to *may* plus (fake) past tense resulting in an unlikelihood inference. Past tense is defined in terms of non-actual veridicality (Karawani 2014, in the spirit of Giannakidou's 1997 non-veridicality notion and Iatridou's 2000 exclusion feature). The unlikelihood inference is argued to be presuppositional.

item (*must*) and the negation of that item's dual (*may not*):

(3) #It must be raining, but it may not be raining.

As we show below, empirical support for this analysis of *may* relative to *might* is clear in one experiment, where speakers were asked directly to assess the strength of *may* and *might*. A complete story of epistemic modality in English - including what the dual of *must* is - will need to reconcile this empirical finding. First, however, we recap the distribution of *may* and *might* in English. As is well known, both *may* and *might* have epistemic readings:

(4) Jack may/might be in the office.

As (4) demonstrates (and following Condoravdi 2002), both *might* and *may* can quantify over the present epistemic (information) state. *May* and *might* are closely related in the sense that the latter is the subjunctive/(fake) past counterpart of the former, yet the inflectional marking of *may* and *might* only obscures the similarity of these lexical items' respective distributions, at least in the epistemic domain. Although *might* seems to be inflected for past tense, it is acceptable in present contexts such as (4); moreover, present-tense *may* is acceptable in past contexts such as (5): contrary to expectations, *may* seems to be licensed in real past tense situations and appears to be able to quantify over past information states. Even in sequence-of-tense constructions (e.g. *said that...*) where one expects *may* to not be licensed, native-speaker intuitions regarding examples such as (6a) and (6b) are unclear: according to some speakers, *may* is slightly dispreferred, but no one we consulted considered it to be completely ungrammatical. Similar things can be said for the case of subjunctive conditionals (7).

(5) I was so sick that I thought that I might/may not make it to school yesterday.

(6) a. He said he might go. / b. (?) He said he may go.

(7) If Mr. Smith were to win the election, he (?)may/might appoint a new sheriff.

However, the respective distributions of *may* and *might* diverge more substantially in non-epistemic contexts. *May*, unlike *might*, is licensed in indicative deontic contexts, (8). *Might*'s distribution as a deontic modal seems to be restricted to question contexts, and semantic intuitions in relation to *may* are crisper here than for the epistemic contexts discussed above. Namely, White (1975) remarks that *may* and *might* are not synonymous in contexts such as (9), presented below: *might he take it?* seems to express "a more tentative request" than does *may he take it?* (1975: 49).

(8) You may/#might go now!

(9) May/might he take it?

Returning to epistemic uses, it appears as though *may* and *might* are not clearly distinguished, at least from the standpoint of grammatical distribution. However, there is a persistent claim in the descriptive literature on English that *may* and *might* are distinguishable on semantic grounds. For example, Leech (2004) claims that "The effect of the hypothetical auxiliary [*might*], with its implication 'contrary to expectation', is to make the expression of possibility more tentative and guarded [relative to *may*]." Similarly, Nuyts (2001: 209) notes that "Obviously *might* used to be the past of *may*, but the past tense has been entirely reinterpreted as an 'epistemic past', i.e. a weakener of the epistemic qualification." This intuition is taken into serious consideration by CKV: in their theory of epistemic modality, *may* is formally stronger than *might*. This feature of their analysis also leads them to revise the 'normal' duality assumption of *might* and *must*.

3 Experiment 1

In our first experiment, we were interested in determining whether speakers' intuitions about the strength of *may* indeed varied relative to strength intuitions for *might*. To create our stimuli, we first gathered data from the Corpus of Contemporary American English (Davies 2008-) where *may* and *might* were employed in epistemic contexts. We identified such 10 sentences, of which 5 contained a *might* phrase while the other 5 contained a *may* phrase. We then modified these sentences to improve discourse coherence. We systematically manipulated these sentences such that they contained either *may*, *might*, *could*, or *must*. The result was 40 sentences: 10 sentences with four possible modal configurations. An example paradigm is presented below:

Confounding factors {*may/might/could/must*} *have skewed the results of the doctors' study.*[2]

We then recruited Amazon Mechanical Turk workers (n = 61) to provide responses to the following two questions about these sentences:

i. *"According to the speaker of the above sentence, how likely is it that* [paraphrase of the prejacent]*?"*

ii. *"How certain is the speaker that* [paraphrase of the prejacent]*?"*[3]

In the vein of a recent experimental investigation into the semantics of epistemic *must* by Scontras et al. (2016), our participants assessed likelihood and certainty by answering the above questions on a slider scale. For the first question, the left end of the scale read "0% likely (impossible)" and the right end read "100% likely (guaranteed to be the case)". For the second question, the left and right ends were "0% certain (certain it is not the case)" and "100% certain (absolutely certain)", respectively. We collected 15 responses for our 40 sentences (600 responses total). We allowed participants to answer for as many sentences as they liked: the mean number of responses per participant was 9.84. Participants were paid $0.05 per response.

3.1 Results and Discussion

	Mean certainty rating	Mean likelihood rating
may	54.71	60.21
might	46.67	54.83
could	46.32	54.97
must	75.49	79.04

The left end of each slider scale was coded "1" in our analysis, while the right end was coded "100". The above table shows the mean response values for our two questions, broken down by modal condition. We performed pairwise linear mixed effects regression analyses using the lme4 package (Bates et al. 2014) in R to predict certainty ratings from a fixed effect of modal as well as by-participant and by-item random intercepts. Certainty ratings were higher in the *may* condition than in the *could* condition ($\beta = 9.07$, $SE = 2.23$, $t = 4.06$, $p < 0.001$). However, there was no difference in certainty ratings between the *might* and *could* conditions ($\beta = 1.29$, $SE = 2.13$, $t = 0.607$, $p < 0.55$). A similar pattern of results held for the likelihood ratings, which were higher in the *may* condition than in the *could* condition ($\beta = 5.91$, $SE = 1.87$, $t = 3.16$, $p < 0.005$); however, there was no difference in likelihood ratings between the *might* and *could* conditions ($\beta = 0.63$, $SE = 1.74$, $t = 0.365$, $p < 0.72$). In this experiment, the difference between *might* and *may* was significant in pairwise comparison tests for likelihood

[2] For some stimuli, we extraposed *must* in order to avoid deontic interpretations, for example:
In order to stay competitive, the company {*may/might/could*} *need to outsource its production.*
It must be the case that in order to stay competitive, the company needs to outsource its production.

[3] Note that we take likelihood to be a property of the sentence while certainty of p is a subjective assessment of speaker's certainty.

ratings ($\beta = -5.34$, $SE = 1.62$, $t = -3.30$, $p < 0.001$) and certainty ratings ($\beta = -8.07$, $SE = 2.08$, $t = -3.89$, $p < 0.005$). Additionally, as shown above, we found evidence that *may* was stronger than another epistemically-construed modal (*could*) where *might* was not, on both the metric of likelihood and the metric of certainty.[4]

We interpret this as evidence of a distinction between *may* and *might*, in the sense that *may* appears to encode higher likelihood of the prejacent as well as a higher degree of speaker certainty of the prejacent. This result is consistent with claims from the descriptive literature on English - discussed above - that epistemic *might* is the 'tentative' or 'qualified' counterpart of epistemic *may*. Furthermore, this result is predicted by the analysis offered by CKV.

4 Experiment 2

The results of Experiment 1 suggested that speakers of English are sensitive to a difference in strength between *may* and *might*. With Experiment 2, we began our investigation of whether and how this distinction is relevant in discourse. For Experiment 2, we took as our point of departure the *Dismissive Agreement* paradigm of Rudin (2015), in which a speaker B may assert *might p* even when *p* is only of marginal possibility given B's epistemic state, as in the example below (Rudin 2015: (1)):

(10) **A:** Paul might come to the party.
B: Yeah, he might, but it's extremely unlikely.

Consistent with CKV's analysis, Rudin (2015) invokes (10) to build the case for a semantics of *might* whereby *might* semantically encodes nonzero likelihood of the prejacent but may give rise to the implicature that "the prejacent is likely enough to be relevant" (Rudin 2015: 596). For our purposes, the key intuition of examples such as (10) is that B's response is a natural way to communicate that for all B knows, the likelihood that Paul comes to the party is extremely low but nonzero. One additional prediction of CKV, however, is that because *may* encodes the 'likely possibility' of *p*, B's response in (11) should be disfluent relative to her response in (10):

(11) **A:** Paul may come to the party.
B: (?)Yeah, he may, but it's extremely unlikely.

To test this, we produced three *Dismissive Agreement* discourse contexts, systematically manipulated such that the featured modal was either *may* or *might*. An example paradigm is shown here: *Context: Bill and Simon are discussing the future of their city's public zoo.*

Bill: *The local zoo {might/may} be shut down by city government this year.*

Simon: *Yeah, it {might/may} be, but it's extremely unlikely.*

The discourses always included an interlocutor "Simon" responding to an interlocutor "Bill". In an online acceptability judgment experiment, we asked workers on MTurk (n = 240, US IP Addresses, Approval Rating > 80%, compensation = $0.15) to "rate the extent to which you think, in the given scenario, that Simon's reply is a natural response to Bill's statement". Participants rated on a scale of 1 to 7, with 1 being "completely unnatural", and 7 being "completely natural". Each participant provided a rating for just one of our six items (three discourse contexts × two modal conditions = six items), and we used the UniqueTurker script (https://uniqueturker.myleott.com/) to prevent multiple participation. We had 240 ratings.

[4] We also coded each of our ten sentence contexts according to whether the sentence originally contained a *may* or a *might* in the COCA. Including this variable as a random or fixed effect did not improve the models for likelihood ratings ($p < 0.67$) or for certainty ratings ($p < 0.77$).

4.1 Results and Discussion

Condition	Average naturalness rating
may	5.72
might	5.62

As the above table shows, naturalness ratings for *may* and *might* were virtually indistinguishable, contrary to what one might expect given the strength asymmetry of *may* and *might* argued for by CKV and suggested by the results of Experiment 1. That is, in contexts where the responding interlocutor Simon wished to assert that the prejacent is an extremely unlikely possibility, *may* appeared to be no worse than *might*.[5]

5 Experiment 3

In Experiment 3, we explored the potential implications of a strength asymmetry between *may* and *might* for semantic duality hypotheses in the epistemic modal domain. First, note that if *may* is indeed stronger than *might*, and if *may* is the dual of *must*, then we are committed to the idea that *must* is weak - because its dual is "stronger" than the existential epistemic possibility modal. Moreover, if we accept CKV's claim that *may* is the dual of a weak *must* and *might* is indeed the bare existential epistemic modal, then we make the following empirical prediction: whereas *must p, but may not p* should always be contradictory given their duality relationship, there may be contexts in which *must p, but might not p* is not contradictory. As a first step towards investigating this, we considered Lassiter (2016), who reports a series of naturally-occurring examples of epistemic *must* in non-maximal certainty contexts, in which the *must* claim is preemptively (12) or subsequently (13) hedged:

(12) I don't know for sure, sweetie, but she must have been very depressed. A person doesn't do something like that lightly. (Lassiter 2016:7, ex. 16)

(13) This spot might be good for fishing, I've always thought, though I haven't seen a soul out there trying. The land must be private, I'm almost certain. (ex. 28)

Lassiter (2016) uses these data to argue, following Karttunen (1972), Veltman (1985) and Kratzer (1991), that *must* is weak. We were particularly interested in these data because they were contexts where a claim of *must p* allowed for the epistemic possibility of *not p*. Thus, they seemed to us to be particularly good contexts to investigate whether *must p* could be felicitously hedged by *might p*, to similar effect as the *I don't know for sure* and *almost certain* hedges of (12) and (13). Because we predict that *must p, but may not p* should always be a contradiction, however, we also predicted that it should be considerably worse than *must p, but might not p* in these same contexts.

For our experiment, we elicited acceptability judgments (MTurk workers, US IP Addresses, Approval Rating > 80%, compensation = \$0.15) of sentences – presented in a 3-4 sentence discourse frame – based on Lassiter's (2016) original examples and modified to achieve our own experimental aims. We tested the acceptability of hedging an assertion of *must p* with *may*

[5] It is possible that a distinction between *may* and *might* in this experiment was obfuscated by the fact that Simon's response always echoed the modal in Bill's statement. Taking this a step further, perhaps the echoic usage of the modal signified something weaker than full agreement with Bill and hence cannot be construed as Simon's full commitment to the modal claim. On this view, *Yeah, Paul may come to the party, but it's extremely unlikely* is potentially not interpeted as a contradiction, as Simon's commitment to the first conjunct is weaker than his commitment to the second.

not p, *might not p*, or *not p*.[6] Judgments were provided on a 1 to 7 acceptability scale - with 1 being "completely acceptable" and 7 being "completely unacceptable". The critical sentence was always bolded, and participants were asked to read the entire discourse frame but to rate just the bolded sentence. An example paradigm is shown below:

I think I've found my dream car at a used car dealership down the road: a beautiful 1964 white Ford Mustang. The body, paint, and suede interior look pristine. I checked the speedometer, and it shows 38,000 miles. ***The mileage must actually be 138,000, but {it might not be / it may not be / it isn't}.*** *At any rate, the car drove beautifully during the test drive!.*

Additionally, for every discourse context, we included two modifications where there was no second conjunct and where the *must* was replaced by either a *may* or *might*. These conditions were included to investigate whether any difference in acceptability between *must p, but may not p* and *must p, but might not p* could be attributed to a default preference for one of the two modals in these discourse contexts.

I think I've found my dream car at a used car dealership down the road: a beautiful 1964 white Ford Mustang. The body, paint, and suede interior look pristine. I checked the speedometer, and it shows 38,000 miles. ***The mileage {may/might} actually be 138,000.*** *At any rate, the car drove beautifully during the test drive!.*

We had three discourse contexts with five manipulations each, for a total of 15 items. Each participant provided a rating of just one item, and we used the UniqueTurker script to prevent multiple participation. We had 40 ratings for each of our 15 items, for a total of 600 responses.

5.1 Results and Discussion

must p, but $\neg p$	*must p, but might* $\neg p$	*must p, but may* $\neg p$	*might p*	*may p*
4.01	4.12	4.36	5.57	5.69

The above table provides the average ratings for the five manipulations across our three discourse contexts.[7] Hedges of *must* with either *might not* or *may not* are more or less equally degraded in our data: we performed pairwise ordinal logistic regression analyses using the MASS package (Venables and Whitney 2002) in R to predict acceptability ratings from a fixed effect of discourse manipulation, and we found that these two aforementioned conditions did not differ significantly ($\beta = 0.23$, $SE = 0.23$, $t = 1.01$, $p < 0.32$). It appears to the extent that *must p, but may not p* is felt to be contradictory, so is *must p, but might not p*. Moreover, the *may p* and *might p* conditions did not differ significantly ($\beta = -0.26$, $SE = 0.24$, $t = -1.12$, $p < 0.27$). Thus, in Experiment 3, we could find no evidence that speakers were distinguishing between *might* and *may* in these discourse contexts: accounts arguing in favour of *might* existing in a unique duality relationship with *must*, must reconcile the data here and offer an account of why *must* and *may* do not exist in a duality relationship.[8]

[6] The *not p* condition – a guaranteed contradiction – was included as a baseline comparison against which to assess the acceptability of the other two continuations.

[7] Two data points were excluded due to the failure of the participant to answer the question.

[8] It is also notable – and perhaps unexpected – that there was no significant difference in acceptability between the *must p, but not p* condition and the *must p, but might not p* condition ($\beta = 0.15$, $SE = 0.228$, $t = 0.663$, $p < 0.51$) also between *must p, but not p* and *must p, but may not p* ($\beta = 0.39$, $SE = 0.228$, $t = 1.69$, $p < 0.1$). This is surprising given that a continuation of *not p* certainly feels 'more' contradictory than a continuation of *might p* in these contexts. Thus, it could be that there is **some** distinction between the *might not p* and *may not p* continuations of *must p* but that our paradigm was not sensitive to finer distinctions of acceptability. In any case, there was no obvious dropoff in acceptability from *may not p* to *might not p*.

6 Discussion - Strength, Duality, and Social Meaning

Though we have found some evidence of a strength asymmetry between *may* and *might*, consistent with the analysis of CKV, we did not find any direct empirical support for the relevance of this asymmetry in contexts of language use: Experiment 2 suggests that *may* is not significantly worse than *might* to express extremely low but nonzero likelihood of the prejacent; Experiment 3 suggests that *must p, but might not p* is not any less disfluent than *must p, but may not p* - at least not in the discourse contexts we explored. How do we reconcile these results with claims from the descriptive literature on English that *might* is the weak epistemic possibility counterpart of *may*? One potential explanation is that *might* and *may* have lost a distinction in epistemic strength as *may* disappears in spoken English. In a corpus study conducted by Bowie et al (2013), it is reported that there has been a significant drop in the use of *must* (-54%) and *may* (-39%) but a rise in the use of *might* (5%). On this story, we might reasonably have had a debate as to what the dual of *must* is several decades ago, when both *might* and *may* were salient (and hence semantically distinguished) members of the spoken lexicon.

There is, however, more to be said here from the diachronic perspective: the shift in the usage of modals in informal spoken registers is a phenomenon that has been observed in languages other than English, including by Gonzales et al. (2017), who report that modals encoding low speaker certainty are generally supplanting markers of high certainty in informal spoken registers of Catalan. A concurrent shift favoring weak epistemic modals is taking place in German, where *könnten* is increasinlgy preferred to *dürften* in everyday discourse.[9] Generally, the trend appears to be a shift away from markers of high certainty. Given this general trend, we straightforwardly account for the rise of *might* and concurrent fall of *may* in English on a semantic account of modals - such as CKV's - whereby *might* is weaker than *may*.[10]

Another story (not incompatible with the diachronic facts mentioned above) makes reference to sociolinguistic theory and to the distributional differences of *might* and *may* across registers of English. Biber et al (1999) report a dichotomy between the use of *may* and *might*: in formal (e.g. academic) registers the use of *may* is significantly higher, as opposed to a higher use of *might* in informal registers. If *may* and *might* are tied to a social register, then we might expect them to encode different **social meaning**: whereas sentence meaning (as analyzed by semanticists) conveys information about the world (facts or thoughts), social meaning (as analyzed by sociolinguists) conveys information about the identity of the speaker, which in turn may enrich sentence meaning. In the following, we will sketch how pragmatic enrichment via social meaning can help us make sense of the empirical picture of *may* and *might*. To begin, consider the following contexts a doctor may find herself in. In both (14) and (15), she makes a statement regarding the sickness afflicting her patient and raises the possibility that it is cancer:

(14) *Doctor, to colleagues:* "This may/might be cancer."

(15) *Doctor, to a patient or patient's family:* "This may/might be cancer."

When asked to compare *may* and *might* sentences in (14) and (15), English native speakers that we consulted were able to report a qualitative difference between the two epistemic possibility modals more readily than in what we saw in Experiments 2 and 3. Namely, in the case of the doctor speaking to her colleagues, native speakers felt that *might* conveyed that the doctor is "not so sure that it is cancer." Intuitions differed in the case of talking to the patient or

[9] We thank Hubert Truckenbrot (p.c.) for pointing this out.

[10] For some opinionated thoughts on the cultural forces driving this diachronic change, consider these selected quotes from a recent New York Times article (April 30, 2016) entitled *Stop Saying: I feel like*: "[T]here is a tendency to commit less [...] a reflex to hedge every statement ."

patient's family: the use of *may* in this context is taken to convey a more aloof, sterile, attitude from the doctor, whereas the use of *might* is taken to involve politeness, empathy, breaking the news softly, etc – there is no perception of uncertainty relative to *may*.

The example illustrates that speakers' intuitions regarding the difference between *may* and *might* depend on social context and the intended social meaning of interlocutors – two factors which we did not systematically investigate in our experiments. Note that the contrast is further reflected when the modals are modified: in the doctor to patient('s family) example, a use of "very well" is particularly marked, for the reason that strengthening the possibility clashes with the doctor's choice for *might*.

(16) *Doctor, to patient('s family)*: "This may/#might very well be cancer."

Speakers learn to access implicit attitudes towards speakers of different linguistic varieties and are competent in shifting between styles to achieve being perceived in one way or another (Eckert 2000; Beltrama 2016; Burnett 2017). A speaker using *may* could risk being perceived as arrogant or pretentious, whereas by using *might* s/he has better chances for coming across as more friendly, sincere, and supportive – even at the risk of being perceived as inarticulate, doubtful, or hesitant. Conversely, a speaker may choose to risk being perceived as authoritative, by using *may*, so as to not be perceived as inarticulate, incompetent, or casual for using *might*. The two possibilities can be seen as pragmatically enriched in the following way:
May: *"speaker is committed to a possibility"*
Might: *"speaker is not committed to a possibility"*

The CKV analysis can explain the contrast detected in (14) and (15) in terms of a core semantic strength asymmetry. Being stronger, *may* introduces a possibility to reckon with, while *might* introduces a possibility that can be taken to be unlikely. Note that this unlikelihood inference is conditioned on social context: one can also use *might* when one means the stronger sense but stops short of greater commitment out of social considerations.[11] In other words, one can use *might* even when one does not wish to communicate that something is unlikely, *per se*. This is a particularly available option when you yourself expect p, but your interlocutor does not expect or is averse to the possibility of p (as in (15)): the decision to use *might* reflects face-saving politeness considerations given the incongruence of epistemic states between interlocutors. In the doctor-to-colleagues *it may be cancer* example in (14), the doctor introduces cancer as a likely possibility, one that needs to be reckoned with, while in the *it might be cancer* example, cancer is inferred to be a (unexpected, unlikley) possibility, or the speaker is perceived as hedging her own certainty or authoritativeness. In this professional interaction between colleagues, there is no reason for the doctor to be polite or sensitive in the same way she would be to a patient's family; thus, *might be cancer* reliably communicates "not so sure it's cancer" with colleagues but not with patients.

Clearly, the respective expressive capabilities of *might* and *may* are at least in part a function of their respective abilities to convey social meaning in a given context. Indeed, the influence of social meaning may well have confounded our attempts to disentangle *may* from *might* in the experiments we report here. Future research must take this context sensitivity seriously. That epistemic modals in English are sensitive to register and social meaning also has implications for the question with which we began the paper: what is the dual of *must*? At a minimum, an answer to this question will need to disentangle the semantics of epistemic modals from their distributional patterns across registers and contexts of use, as well as address the fact that *might* appears to be replacing *may* in spoken English. Indeed, the final answer may prove

[11] See Krifka (2015, et seq) for a formal account of commitment.

to be variable depending on the social dynamics of a given discourse context. We leave this exploration of the interaction between social meaning and modality to future work.

Bates, D., Mächler, M., Bolker, B., & Walker, S. 2014. Fitting linear mixed-effects models using lme4. Journal of Statistical Software, 67(1), 1-48.
Beltrama, A. 2016. Bridging the gap: intensifiers between semantic and social meaning. PhD dissertation. University of Chicago.
Biber, D., S. Johansson, G. Leech, S. Conrad, & E. Finegan. 1999. The Longman grammar of spoken and written English. London: Longman.
Bowie, J, Wallis, S, Aarts, S. 2013. Contemporary change in modal usage in spoken British English: mapping the impact of "genre". In: M. Carretero, J. Arús Hita, J. Van der Auwera, JI. Marín-Arrese (eds.) English modality: core, periphery and evidentiality. 57-94. De Gruyter: Berlin and New York.
Burnett, H. 2017. Sociolinguistic Interaction and Identity Construction: The View from Game-Theoretic Pragmatics. Journal of Sociolinguistics, 21, 2.
Crespo, I., H. Karawani, and F. Veltman. 2017. Expressing expectations. In D. Ball and B. Rabern (eds.), The Science of Meaning: Essays on the Metatheory of Natural Language Semantics. Oxford Univerisity Press.
Davies, Mark. 2008. The Corpus of Contemporary American English (COCA): 520 million words, 1990-present.
Degen, J. 2015. Investigating the distribution of some (but not all) implicatures using corpora and web-based methods. Semantics and Pragmatics 8(11). 1-55.
Degen, J., Scontras, G., Trotzke, A., & Wittenberg, E. 2016. Definitely, maybe: Approaching speaker commitment experimentally. MS, Stanford University.
Eckert, P. 2000. Language variation as social practice: The linguistic construction of identity in Belten High. Wiley-Blackwell.
von Fintel, K. & A. Gillies. 2010. Must...stay...strong! Natural Language Semantics 18(4), 351–383.
Giannakidou, A. 1997. The Landscape of Polarity Items. Ph.D. thesis. U Groningen.
Gonzales, M., Roseano, P., Borras-Comes, J., Prieto, P., (2017). Epistemic and evidential marking in discourse: Effects of register and debatability. Lingua 186-187. 68-87.
Iatridou S. 2000. The grammatical ingredients of counterfactuality Linguistic Inquiry. 31: 231-270.
Karawani, H. 2014. The Real, the Fake, and the Fake Fake in Counterfactual Conditionals, Crosslinguistically. Doctoral dissertation, University of Amsterdam. LOT Dissertation Series 357, 2014.
Karttunen, L. 1972. Possible and must. In John Kimball (ed.), Syntax and semantics, vol. 1, Seminar Press.
Kratzer, A. 1991. Modality. In Arnim von Stechow and Dieter Wunderlich (eds.), Semantics: An international handbook of contemporary research. 639–650. de Gruyter.
Lassiter, D. 2016. Must, knowledge, & (in)directness. Natural Language Semantics 24(2). 117-163.
Leech, G. N. 2004. Meaning and the English verb. Pearson Education.
Krifka, M. 2015. Bias in Commitment Space Semantics: Declarative questions, negated questions, and question tags. Semantics and Linguistic Theory (SALT) 25, 328-345.
Lambert, W.E., R.C. Hodgson, R.C. Gardner, and S. Fillenbaum. 1960. Evaluational reactions to spoken language. Journal of Abnormal and Social Psychology 60.1: 44-51.
Nuyts, J. 2001. Epistemic modality, language, and conceptualization: A cognitive-pragmatic perspective. John Benjamins Publishing.
R Core Team 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.
Rudin, D. 2015. Deriving a variable-strength might. Sinn und Bedeutung 20. 587–603.
Veltman, F. 1985. Logics for Conditionals. Ph.D. thesis, University of Amsterdam.
Veltman, F. 1986. Data semantics and the pragmatics of indicative conditionals. In E. Traugott et al. (eds.), On conditionals. 147–167. Cambridge University Press.
Venables, W. N. & Ripley, B. D. 2002 Modern Applied Statistics with S. Fourth Edition.
White, A. R. 1975. Modal thinking. Cornell University Press, Ithaca, N.Y.
Willer, M. 2013. Dynamics of epistemic modality. Philosophical Review 122(1). 45-92.

Explaining the Ambiguity of Past-Under-Past Embeddings*

Carina Kauf[1] and Hedde Zeijlstra[2]

[1] Georg-August-Universität Göttingen, Germany
carina.kauf@uni-goettingen.de
[2] Georg-August-Universität Göttingen, Germany
hzeijls@uni-goettingen.de

Abstract

Past-under-past embeddings have two readings, a simultaneous and a backward-shifted one. While existing accounts derive these readings via distinct mechanisms, be it by means of an ambiguity at the level of LF or via blocking of a cessation implicature, we propose an alternative account which avoids such ambiguity. For us, the meaning of a past tense morpheme, like *-ed*, is comprised of two components. Syntactically, every past tense morpheme carries an uninterpretable past feature [uPAST], to be checked by a (single) covert past tense operator *Op-PAST* carrying an interpretable feature [iPAST]. Semantically, the past tense marker encodes a relative non-future with respect to its closest c-commanding tense node (informally: 'not later than'), immediately yielding the two distinct readings.

1 Introduction

Constructions in which a past tense is embedded under a matrix past tense have two readings: a simultaneous reading and a backward-shifted one.

(1) John said Mary was ill.
   a. John, at some $t' <$ utterance time, t_u : "Mary is ill." (simultaneous reading)
   b. John, at some $t' < t_u$: "Mary was ill." (backward-shifted reading)

The availability of the simultaneous reading for past-under-past sentences – commonly referred to as Sequence of Tense (= SoT) – has been a prevalent topic of research for several years. Drawing on the discussion of whether past is an absolutive or a relative tense, the extraordinarity of the SoT phenomenon can be explained in the following way: In Reichenbachian [Rei47] terms, past morphology establishes a set, absolutive relation between the sentence's utterance time (UT), its event time (ET), and its reference time (RT): ET = RT < UT. In clauses in which a past tense morpheme is embedded under a matrix past, we thus predict one of the following configurations: Either both past tense morphemes establish the said relation between ET, RT, and UT independently of each other, yielding a free order of the two past ETs with reference to each other, or the matrix morpheme's ET locally changes the variable of the embedded morpheme's UT to a time in the past, leading to an obligatory backward-shift. Whereas the former option generates too many possible readings, the latter option generates too few; it can neither explain the existence nor the empirical prominence of the simultaneous reading. Hence, an additional factor must be at play in SoT cases.

In all existing accounts, the simultaneous and the backward-shifted readings are derived via distinct mechanisms. Most implement the distinction as an ambiguity at the level of LF, assuming either a syntactic rule of tense deletion under certain conditions [Ogi95, vS95], a zero

*This paper greatly benefitted from discussions with Cleo Condoravdi.

tense in the embedded clause [Kra98], a feature transfer mechanism that transmits temporal relations [Abu97], or a combination of the last two [GvS10], among others. However, the systematic, cross-linguistic availability of this ambiguity casts doubt on whether it should indeed be attributed to two different LFs, instead of receiving a more principled explanation. A different type of proposal therefore assumes the ambiguity to take place at a higher level, i.e. pragmatics, and derives the two distinctive readings on the grounds of the blocking/existence of a cessation implicature [Alt16, AS12].

2 Proposal

In this paper, we propose an alternative approach for SoT that avoids ambiguity. For us, the meaning of a past tense morpheme, like *-ed*, is comprised of two components. In line with many others (e.g. [vS03, Sto07, Zei12]) we assume that syntactically, each past tense morpheme carries an uninterpretable past feature [uPAST], to be checked by a covert past tense operator *Op-PAST* that carries an interpretable feature [iPAST]. (In this paper we restrict ourselves to temporal interpretations of past tense morphology, taking non-past, non-factual readings of past tense morphology out of consideration.) We implement this past tense operator in the following way:

(2) $[\![\textit{Op-PAST}]\!] = [\ \lambda t^*.\ \lambda P.\ \exists t < t^*\ \&\ P(t)\ ]$

At matrix level, t^* in principle applies to t_u and for the sake of easiness we will take *Op-PAST* to denote $[\ \lambda P.\ \exists t < t_u\ \&\ P(t)\ ]$ in these cases. Nevertheless, later in this paper we will discuss examples in which the value deviates from the default, providing evidence for the necessity of the more complex definition of the operator given in (2).

That past tense takes higher scope than the surface position of the past tense marker has been well-established in the literature [Kle94, Ogi96, Abu97, vS02, Zei12]. Evidence for this comes from examples like:

(3) Wolfgang played tennis on every Sunday. [vS06]

The interpretation of (3) is one in which past tense outscopes the distributive quantifier *every Sunday*, yielding the paraphrase in (4-a). If the scopal order is reversed, the only possible interpretations are (4-b) and (4-c), both of which are infeasible (cf. [Zei12, vS02, vS05]).

(4) (In)feasible paraphrases of (3) as the result of different scopal orderings
- a. = 'For every Sunday in the past, there is a time t at which Wolfgang plays tennis.'
- b. ≠ 'There is past time on every Sunday at which Wolfgang plays tennis.'
- c. ≠ 'For every Sunday, there is time before it such that Wolfgang plays tennis at that time.' [vS06]

Even though the locus of past tense is different from its overt instantiation, i.e. the tense marker *-ed*, this does not entail that the past tense morpheme is semantically vacuous. In fact, our approach deviates from existing agreement accounts in assuming that both the covert operator and the past tense morpheme are semantically active. We argue that the tense marker encodes a relative non-future with respect to its closest c-commanding tense node (informally: 'not later than') (cf. [Kra98, vS03]). Accordingly, it has the following denotation:

(5) $[\![\text{-ed}]\!] = [\ \lambda t.\ \lambda P.\ \exists t'.\ t' \leq t\ \&\ P(t')\ ]$

In this context, the expression $t' \leq t$ is defined to mean that the lower boundary of the time interval t' is not later than the lower boundary of the time interval t. Hence, an event happening at time t' starts either strictly earlier than or at the same time as an event happening at time t, but can never start later than it.

A simple sentence such as *Susan left* thus receives the following interpretation:

(6) Susan left.

a. [*Op-PAST*$_{[\text{iPAST}]}$ [Susan leave-ed$_{[\text{uPAST}]}$]]
$\exists t' < t_u$ $\exists t^2 \leq t'$

b. $\exists t' < t_u$ & [$\exists t^2 \leq t'$ & leave(Susan, t^2)]

c. *There is a time t' strictly before the utterance time t_u and Susan leaves at a time no later than t'.*

The analysis deviates from standard analyses in that it introduces an ambiguity with respect to the ordering of t' and t^2: they either refer to the same point in time or the latter precedes the former. Nevertheless, for unembedded sentences such as (6), this ambiguity remains indistinguishable. As a result, the meaning of (6) remains the same, i.e. the time of Susan's leaving is strictly prior to the utterance time.

Whereas the 'not later than'-ambiguity yielded by the semantic meaning component of past tense morphology (i.e. (5)) does not change the meaning of unembedded sentences, it directly entails that every past tense embedded under another past tense is ambiguous between a simultaneous and a backward-shifted reading. This becomes evident from the hierarchy of the tense nodes our proposal yields for such cases: The absolutive past tense operator places the sentence prior to the utterance time, thus providing the temporal head of the time chain; all other past tense nodes semantically express a relative non-future with respect to their closest c-commanding tense node. In cases where *Op-PAST* can check all [uPAST] features via multiple agree, this configuration yields a totally ordered set of tense nodes from the matrix past operator to the most embedded past tense node. For an illustration of our proposal, consider the following derivation (where we take *say* to be an extensional predicate):

(7) John said that Mary was ill.

a. [*Op-PAST*$_{[\text{iPAST}]}$ [John [say-ed$_{[\text{uPAST}]}$ [that [Mary [be-ed$_{[\text{uPAST}]}$ ill.]]]]]]
$\exists t' < t_u$ $\exists t^2 \leq t'$ $\exists t^3 \leq t^2$

b. $\exists t' < t_u$ & [$\exists t^2 \leq t'$ & say(John, t^2, [$\exists t^3 \leq t^2$ & be-ill(Mary, t^3)])]

c. *John's saying is strictly before the utterance time t_u and Mary's being ill starts out no later than the time as John's saying.*

The covert past tense operator in (7) places the proposition at some time $t' < t_u$. Since both of the embedded past tenses lie within the syntactic domain of this higher operator, it can check both morphemes' [uPAST] features via multiple agree. This gives rise to the following semantic relations: The two past tense morphemes each introduce a relation of relative non-future with respect to their closest c-commanding tense node. Hence, t^2 is interpreted as a relative non-future with respect to t', and t^3 constitutes a relative non-future with respect to t^2. The backward-shifted reading of (7) then arises in case that $t^3 < t^2$, while the simultaneous interpretation is yielded for $t^3 = t^2$.

One may wonder why a second *Op-PAST* operator may not have been included in the embedded clause to check off the lower past tense morpheme's [uPAST] feature. Zeijlstra [Zei12] proposes that the number of operators is regulated by economy principles: A second operator may only be included when neccessary. Since the covert past tense operator in (2) can check all

of the uninterpretable past tense features in its syntactic domain (via multiple agree), multiple past tense morphemes in principle require the presence of only one past tense operator. With respect to Zeijlstra's economy constraint, this means that when one *Op-PAST* can check all present [uPAST] features, no further *Op-PAST* may be included.

3 Explaining Challenging Past-Under-Past Embeddings

The previous section has shown that for standard SoT sentences, our theory yields the correct results: it derives the simultaneous and the backward-shifted reading in past-under-past embeddings, but crucially not the forward-shifted reading. Nevertheless, in principle, any theory of SoT also has to account for more complex cases of temporal embeddings, e.g. cases in which an embedded past tense refers to a time in the future, or in which a past tense is not even ordered relatively to a higher tense node. This section is devoted to showing how our approach deals with these cases of more challenging SoT sentences.

3.1 Complement Clausal Embeddings

3.1.1 Future Reference in Past-Under-Past Embeddings – Embedded Past

The following examples show that a past-embedded past tense can make reference to a time interval that lies strictly after the time of utterance:

(8) He said he would buy a fish that was still alive.

(9) He decided a week ago that in ten days he would say to his mother that they were having their last meal together.

The challenge comes from the interpretation of their most embedded past tense forms (underlined in the above examples). As has been well-established in the literature (cf. e.g. [Abu97, Ogi95]) the most prominent reading of these past tense morphemes is one of simultaneity with respect to their c-commanding tense nodes, which, in these examples, have been shifted to a time later than the matrix time by means of the modal *would*. In (8) a past tense can thus be used to describe the state of a fish's being alive, even if this time interval lies in the strict future of the utterance time. In (9), past tense morphology is used to describe a containment relation between the time of the saying event (which is necessarily interpreted as being later than the matrix time due to the overt modifier clauses) and the event time of the last meal. In this example, it is further the case that the backward-shifted relation between the most embedded past tense and its c-commanding tense node that our analysis predicts to be available is independently blocked due to additional aspectual information (i.e. the imperfective aspect on *having*); hence, this does not provide a problem for the proposed analysis.

Our approach successfully captures the multiple interpretations of such 'fish-sentences' under the assumption that *would* is a combination of the operator *woll* (a tense operator that places the evaluation time of a proposition in the relative future of the sentence's current evaluation time) plus a [uPAST] feature that restricts it to past tense sentences (again taking non-past, non-factual readings of *would* out of consideration here).

(10) $[\![\text{woll}_{[\text{uPAST}]}]\!] = [\ \lambda t.\ \lambda P.\ \exists t'.\ t' > t\ \&\ P(t')\ ]$[1]

[1] Here, we ignore the modal contribution of the operator *woll* in terms of universal quantification over possible worlds (cf. e.g. [Ipp13]), which is orthogonal to the analysis presented in this paper.

Crucially, we thus assume that *would* does not carry complete past tense morphology, which is why it does not introduce a relative-non future relation with respect to its closest c-commanding tense node.

(11) John said he would buy a fish that was alive.

a. [*Op-PAST*$_{[\text{iPAST}]}$ [John [say-ed$_{[\text{uPAST}]}$ [he [woll$_{[\text{uPAST}]}$ [buy a fish [that
$\exists t' < t_u$ $\exists t^2 \leq t'$ $\exists t^3 > t^2$
be-ed$_{[\text{uPAST}]}$ alive.]]]]]]]
$\exists t^4 \leq t^3$

b. $\exists x$ [fish(x) & $\exists t' < t_u$. $\exists t^2 \leq t'$: say(John, t^2, [$\exists t^3 > t^2$: buy(he, t^3, x) & $\exists t^4 \leq t^3$: alive(x, t^4)])]

c. *There is a time t^4 which is the time of a contextually salient fish's being alive, and t^4 is prior or equal to some time t^3. The time t^3 is the time of John's buying the fish which lies strictly after t^2, i.e. the time of John's saying event. t^2 is prior or equal to t' which, in turn, is a time strictly before the utterance time t_u.*

What is essential about the analysis is that the most embedded past, i.e. *was* in (11), is ordered prior or equal to the time of the buying, and not prior to any other time, such as the matrix time or the utterance time. This correctly yields a later-than-matrix interpretation of the embedded past tense. The simultaneous and the backward-shifted reading of sentence (11) can be made evident via adding suitable modifier clauses:

(12) The two readings of (11):

a. *Backward-shifted interpretation*
John said (three days ago) that (in ten days) he would buy a fish that was alive (a day before).

b. *Simultaneous interpretation*
John said (three days ago) that (in ten days) he would buy a fish that was (still) alive (then).

Similarly, the same applies to example (9). Even when neglecting the temporal modifier clauses, which indubitably place the time of the meal in the future, the formula derived from the tense nodes within the sentence already shows that the time of the meal is not restricted to a past interval. As it is ordered relatively to the future-shifted time of the saying event, the time of the meal can lie strictly after t_u.

(13) He decided (a week ago) that (in ten days) he would say to his mother that they <u>were</u> having their last meal together.

a. [*Op-PAST*$_{[\text{iPAST}]}$ [He [decide-ed$_{[\text{uPAST}]}$ [he [woll$_{[\text{uPAST}]}$ [say to his mother [that
$\exists t' < t_u$ $\exists t^2 \leq t'$ $\exists t^3 > t^2$
they be-ed$_{[\text{uPAST}]}$ having their last meal together.]]]]]]]
$\exists t^4 \leq t^3$

b. $\exists t' < t_u$ & [$\exists t^2 \leq t'$ & decide(he, t^2, [$\exists t^3 > t^2$ & say-to-mom(he, t^3, [$\exists t^4 \leq t^3$ & be-having(they, last meal together), t^4])]]

c. *There is a time t^4 which is the time of their last meal, and t^4 is prior or equal to some time t^3. The time t^3 is the time of his saying and lies strictly after t^2, i.e. the time of his deciding. t^2 is prior or equal to t' which, in turn, is a time strictly before the utterance time t_u.*

3.1.2 Future Reference in Past-Under-Past Embeddings – Matrix Past

Another set of challenging data which invokes a future reference for a past-embedded past tense in a complement sentence is comprised of sentences like the following:

(14) He hoped she tried to kill him first. [Kle16]

The novel challenge posed by these examples lies in the fact that they have an interpretation akin to that of (8) and (9), even though they do not contain an overt future shifter, like *woll*. Klecha [Kle16] argues that the availability of such an independent future-shifted interpretation is restricted to predicates that already have an inherent future orientation, such as *hope* or *pray*. Such a view is in accordance with our proposal; even though the past tense morphology on *hope* places the time of the matrix sentence prior to the utterance time, as a future-oriented predicate, *hope* by itself can shift the evaluation time of its complement proposition to a future point in time – even in the absence of the modal *woll*. Consequently, our analysis derives the correct meaning of these sentences *mutatis mutandis*: Since the forward-shifted evaluation time t' is introduced in the matrix clause (which can lie strictly after the time of utterance t_u), the verb *tried* then simply means *tried at time* t^2, where t^2 is no later than t' (but can also lie in the strict future of t_u).

3.1.3 Future Reference in Past-Under-Future Embeddings – Embedded Past

Future-embedded past tenses (cf. (15)) also give rise to a reading in which the past-marked predicate may take place after the utterance time and thus pose a further challenge to SoT theories.

(15) Alan will think everyone hid.

Since the modal *woll* is instantiated as *will* in this case, it becomes immediately evident that the *Op-PAST* cannot take higher scope than the modal (as otherwise it would be spelled out as *would*). Hence, the underlying structure of (15) must be the following:

(16) [will think [*Op-PAST*$_{[\text{iPAST}]}$ [everyone hide-ed$_{[\text{uPAST}]}$]]]

Taking the same denotation for the modal operator *woll* as in (10), modulo the [uPAST] feature, yields the semantics for *will*. As before, the open value t^* again gets valued by its default value t_u in this case. With *will* taking scope over the past tense operator and changing the local evaluation time t^* against which the past operator gets valued to a time in the future, the correct interpretation of (15) is yielded.

(17) Alan will think everyone hid.

a. [will think [*Op-PAST*$_{[\text{iPAST}]}$ [everyone hide-ed$_{[\text{uPAST}]}$]]]
   $\exists t' > t_u$ $\exists t^2 < t'$ $\exists t^3 \leq t^2$

b. $\exists t' > t_u$ & think(Alan, t', [$\exists t^2 < t'$ & $\exists t^3 \leq t^2$ & hide(everyone, t^3)])

c. *There is a time* t' *in the strict future of* t_u *and Alan thinks at* t' *that there is a time* t^2 *earlier than* t' *such that everyone from a contextually salient group hid at a point* t^3 *no later than* t^2.

Note that if the operator *Op-PAST* entailed an absolutive past ordering of the sentence it takes scope over with respect to the utterance time, these cases could not be accounted for by our proposal. However, as seen in (2), the relation 'prior to time of utterance' is not cooked into the semantics of *Op-PAST*; Instead, the operator is defined as a relative past with respect

to a time variable t^*, whose value may be t_u, but which can also refer to a time interval later than t_u if introduced by an independent source, e.g. by the modal operator *will* (cf. (17)).

3.2 Relative Clausal Embeddings

A further set of data where past-under-past morphology exhibits a deviant behavior from the default is comprised of (non-restrictive) relative clauses: In non-restrictive relative clauses as in example (18), the embedded past can yield any of the following readings: a backward-shifted, a simultaneous and a later-than-matrix one. By contrast, in (19) with a restrictive relative clauses, the later-than-matrix reading is not available [Hei94, Ogi95, Sto07]:

(18) Mary met a woman who was president. [non-restrictive]
   a. In 2000, Mary met a woman who was president in 1995.
   b. In 2000, Mary met a woman who was president in 2000.
   c. In 2000, Mary met a woman who was president in 2004.

(19) Mary was looking for a woman who was president. [restrictive]
   a. In 2000, Mary was looking for a woman who was president in 1995.
   b. In 2000, Mary was looking for a woman who was president in 2000.
   c. *In 2000, Mary was looking for a woman who was president in 2004.

In (18), under the most salient reading, the relative clause is non-restrictive. Example (19) is structurally ambiguous between the relative clause being restrictive and being non-restrictive. The *de dicto* reading is only available, however, in a restrictive relative clause.

Following Enç's [Enç87] observation that relative clause tenses differ from complement clause tenses in allowing an independent, or absolute interpretation, Abusch [Abu88] showed that this only applies to relative clauses that receive a *de re* interpretation (see also [Ogi89, Ogi96]). The *de re*/*de dicto* distinction is strongly connected to the distinction between restrictive and non-restrictive (or appositive) relative clauses, as can also be witnessed in the above examples: (18) contains a non-restrictive and a *de re*-interpreted relative clause. By contrast, under the triggered *de dicto* reading of (19), the relative clause is understood to be restrictive. *De dicto* interpretations are only available in restrictive relative clauses. This enables us to connect the availability of an absolute tense interpretation to the syntactic difference between restrictive and non-restrictive relative clauses.

As is well known, restrictive and non-restrictive relative clauses behave differently with respect to syntactic locality. Whereas non-restrictive relative clauses are syntactically opaque (cf. [Saf86, Fab90, Dem91, Bor92, Arn07] for different accounts for the locality effects of non-restrictive relative clauses), restrictive relative clauses are more accessible. That allows us to entertain the hypothesis (in line with Stowell [Sto07] , though also substantially different) that the past tense morpheme inside a relative clause can have its [uPAST] feature checked against a higher covert tense operator carrying [iPAST], but that the past tense morpheme inside a non-restrictive relative clause cannot do so. Consequently, the latter requires a covert past tense operator of its own, with t^* being valued for the time of utterance. Therefore, a restrictive relative clause allows only a simultaneous reading and a backward shift (when containing past tense morphology embedded by a higher past tense clause), whereas a non-restrictive relative clause in the same situation yields a simultaneous reading, a backward shift and a forward shift. This explains why the two past tense markers in (20) need to be evaluated independently of each other with respect to the time of utterance: Given the syntactic opacity of non-relative clauses, the lower *Op-PAST* in (20) cannot be bound by any higher tense variable, licensing its

existence under the economy principle [Zei12].

(20) Mary met a woman who was president.

a. [*Op-PAST*$_{[iPAST]}$ [Mary meet-ed$_{[uPAST]}$ a woman [who [*Op-PAST*$_{[iPAST]}$ [be-ed$_{[uPAST]}$ president]]]]]
$\exists t' < t_u$ $\exists t^2 \leq t'$ $\exists t'' < t_u$ $\exists t^3 \leq t''$

b. $\exists x$ [woman(x) & $\exists t' < t_u$. $\exists t^2 \leq t'$: meet(Mary, x, t^2) & $\exists t'' < t_u$. $\exists t^3 \leq t''$: president(x, t^3)]

c. *There is a woman x and at t^2, prior or equal to t' which, in turn, is a time strictly before the utterance time t_u, Mary met x, and at t^3, prior or equal to t'' which, in turn, is a time strictly before the utterance time t_u, x is president.*

A restrictive relative clause as (19), for which agreement inside of the relative clause is possible, on the other hand, receives the following interpretation; The most embedded past tense is ordered with respect to the matrix tense and cannot independently be placed prior to the utterance time:

(21) Mary was looking for a woman who was president

a. [*Op-PAST*$_{[iPAST]}$ [Mary be-ed$_{[uPAST]}$ looking for a woman [who [be-ed$_{[uPAST]}$ president]]]]
$\exists t' < t_u$ $\exists t^2 \leq t'$ $\exists t^3 \leq t^2$

b. $\exists x$ [woman(x) & $\exists t' < t_u$. $\exists t^2 \leq t'$: be-looking-for(Mary, x, t^2) & $\exists t^3 \leq t^2$: be-president(x, t^3)]

c. *There is a woman x and at t^2, prior or equal to t' which, in turn, is a time strictly before the utterance time t_u, Mary is looking for x, and at t^3, prior or equal to t^2, x is president.*

4 Advantages of this SoT Approach

The account proposed in this paper has several advantages over existing approaches. First, we do not have to postulate that there is a difference between a real past and a surface past, which is, in fact, a present tense in disguise (cf. e.g. [Ros67, Abu88]). By defining past as non-future, the proposed approach can account for the same cases as the present-in-disguise proposals while at the same time avoiding unwanted ambiguity and instead retaining a clear 1:1 mapping between temporal form and temporal meaning.

Secondly, the proposed SoT account neither is dependent on the intensional/extensional distinction for the embedding predicates [Abu97], nor is it dependent on the stative/eventive distinction of the past embedded verbs [Alt16, AS12]. In her pioneering proposal, Abusch [Abu97] assumes that feature transmission (leading to the SoT effect) only arises with intensional embeddings – a claim that appears to be too strong. As has been illustrated in (7), the analysis we propose in principle also applies to extensional embeddings, yielding the same SoT effects as for intensional embeddings. Existing pragmatic theories of SoT [Alt16, AS12] on the other hand assume that past-under-past embeddings are not ambiguous between a simultaneous and a backward-shifted reading but that they always have a backward-shifted interpretation. According to their theory, the perception of simultaneity arises in the absence of a cessation implicature, which arises only when there is competition with a present tense. Such an absence

of cessation is only licensed by the temporal profile of stative (and not eventive) predicates (i.e. For any tenseless stative clause ϕ, if a moment m is in $[\![\,\phi\,]\!]$, then there is a moment m' preceding m and a moment m'' following m such that m' and m'' are in $[\![\,\phi\,]\!]$) [Alt16, AS12]. Since the absence of cessation is restricted to stative predicates, this type of proposal in principle predicts that past-under-past embedded eventive predicates are always interpreted in a backward-shifted manner, which is the standard assumption. This claim, however, has been refuted by Kusumoto [Kus99], who – with Partee (p.c. to Kusumoto) (cf. [Kho07]) – argues that the examples in (22) have a simultaneous reading even though they embed a past eventive verb:

(22) a. Elliott observed/noticed/perceived that Josephine *got* hurt.
b. He didn't realize that his car *hit* the curb.
c. The pilot was sure that the plane *landed* in the correct spot. [Kus99]

Lastly, the account proposed in this paper is built on a number of parameters (e.g. the no-later-than semantics of past tense morphemes, *Op-PAST* being a relative past operator, a.o.), which, taken together, yields our analysis of past-under-past embeddings. The existence of such parameters opens up a space for variation, which in principle should account for cross-linguistic differences attested with respect to SoT: Whereas English-like SoT-languages may for example encode a constraint in order to rule out an unwanted forward shift of default past-under-past clausal embeddings (e.g. Mary's illness in (7) cannot have started later than at t^2) directly as part of their semantics (cf. (5)), non-English-like SoT languages may exhibit different parameter configurations. An example of a language which is, presumably, built on a different parameter setting is Japanese, for which a past-embedded past tense can only yield a backward-shifted reading (cf. e.g. [Ogi89, Ogi96]). Hence, at least the semantic contribution of Japanese past morphology differs from that of English. A proper investigation of this hypothesis is part of future research.

5 Conclusion

In this paper, we provide a novel SoT-account which avoids ambiguity at the level of LF while at the same time retaining the possibility for both a simultaneous and a backward-shifted reading independent of the temporal profile of the embedded predicates. The two readings are licensed via the weak precedence relation introduced by the semantic meaning component of past tense morphology (i.e. 'no later than' rather than 'strictly earlier than' semantics). We show that this approach, even though various questions are still open, can deal with the same challenges as other SoT approaches and has certain additional advantages as well.

References

[Abu88] D. Abusch. Sequence of tense, intensionality and scope. In *Proceedings of the 7th West Coast Conference on Formal Linguistics*, pages 1–14. CSLI, Stanford, 1988.

[Abu97] D. Abusch. Sequence of tense and temporal de re. *Linguistics and Philosophy*, 20(1):1–50, 1997.

[Alt16] D. Altshuler. *Events, States and Times: An essay on narrative discourse in English.* de Gruyter GmbH & Co KG, 2016.

[Arn07] D. Arnold. Non-restrictive relatives are not orphans. *Journal of Linguistics*, 43(2):271–309, 2007.

[AS12] D. Altshuler and R. Schwarzschild. Moment of change, cessation implicatures and simultaneous readings. In *Proceedings of Sinn und Bedeutung*, volume 17, pages 45–62, 2012.

[Bor92] R. D. Borsley. More on the difference between English restrictive and non-restrictive relative clauses. *Journal of Linguistics*, 28(1):139–148, 1992.

[Dem91] H. Demirdache. *Resumptive chains in restrictive relatives, appositives, and dislocation structures*. PhD thesis, Massachusetts Institute of Technology, 1991.

[Enç87] M. Enç. Anchoring conditions for tense. *Linguistic Inquiry*, pages 633–657, 1987.

[Fab90] N. Fabb. The difference between English restrictive and nonrestrictive relative clauses. *Journal of Linguistics*, 26(1):57–77, 1990.

[GvS10] A. Grønn and A. von Stechow. Complement tense in contrast: the SOT parameter in Russian and English. *Oslo Studies in Language*, 2(1), 2010.

[Hei94] I. Heim. Comments on Abusch's theory of tense. *Ellipsis, tense and questions*, pages 143–170, 1994.

[Ipp13] M. Ippolito. *Subjunctive conditionals: a linguistic analysis*, volume 65. MIT Press, 2013.

[Kho07] O. Khomitsevich. *Dependencies across phases: From sequence of tense to restrictions on movement*. Netherlands Graduate School of Linguistics, 2007.

[Kle94] W. Klein. *Time in language*. Psychology Press, 1994.

[Kle16] P. Klecha. Modality and embedded temporal operators. *Semantics and Pragmatics*, 9:9–1, 2016.

[Kra98] A. Kratzer. More structural analogies between pronouns and tenses. In *Proceedings of SALT VIII. (Cornell Working Papers in Linguistics)*, volume 8, pages 92–110, 1998.

[Kus99] K. Kusumoto. Tense in embedded contexts. 1999.

[Ogi89] T. Ogihara. Temporal reference in English and Japanese. Master's thesis, Doctoral dissertation, University of Texas, Austin, 1989.

[Ogi95] T. Ogihara. The semantics of tense in embedded clauses. *Linguistic Inquiry*, pages 663–679, 1995.

[Ogi96] T. Ogihara. Tense, scope and attitude ascription, 1996.

[Rei47] H. Reichenbach. The tenses of verbs. *Time: From Concept to Narrative Construct: a Reader*, 1947.

[Ros67] J. R. Ross. Constraints on variables in syntax. 1967.

[Saf86] K. Safir. Relative clauses in a theory of binding and levels. *Linguistic Inquiry*, pages 663–689, 1986.

[Sto07] T. Stowell. The syntactic expression of tense. *Lingua*, 117(2):437–463, 2007.

[vS95] A. von Stechow. On the proper treatment of tense. In *Proceedings of SALT V. (Cornell Working Papers in Linguistics)*, volume 5, pages 362–386, 1995.

[vS02] A. von Stechow. Temporal prepositional phrases with quantifiers: Some additions to Pratt and Francez (2001). *Linguistics and Philosophy*, 25(5):755–800, 2002.

[vS03] A. von Stechow. Feature deletion under semantic binding: Tense, person, and mood under verbal quantifiers. In *Proceedings of NELS 33*, volume 33, pages 379–404, 2003.

[vS05] A. von Stechow. Semantisches und morphologisches Tempus: Zur temporalen Orientierung von Einstellungen und Modalen. *Neue Beiträge zur Germanistik*, 4(2):9–54, 2005.

[vS06] A. von Stechow. Types of iF/uF Agreement. Ms. Universität Tübingen, 2006.

[Zei12] H. Zeijlstra. There is only one way to agree. *The Linguistic Review*, 29:491–539, 2012.

Sobel Sequences – Relevancy or Imprecision?*

David Krassnig

University of Konstanz, Konstanz, Baden-Württemberg, Germany
david.krassnig@uni-konstanz.de

Abstract

Sobel sequences were recently split into two independent phenomena by Klecha [5, 6]: Reversible True Sobel sequences and irreversible Lewis sequences. In this paper we show that Klecha's prediction of unidirectionality for Lewis sequences is too strong. To this effect, we propose an alternate analysis, using Lewis' [13, 14] contextualist relevancy-based framework for conditionals, from which a weaker version of Klecha's analysis follows naturally, if we accept Bennett [2] and Arregui's [1] view on how causality affects world similarity. In doing so, we automatically provide an explanation for infelicitous reverse True Sobel sequences, which is, as we also show, a problem for Klecha's current account. Finally, we reunify the analysis of both sequence types under a single overarching linguistic phenomenon by treating the individual sequence types as proper subsets of Sobel sequences.

1 Introduction

For fifty years, starting with the work of Stalnaker [16] and Lewis [11], *Sobel sequences* have played an important role in the debate between strict and variable-strict conditional semantics. A Sobel sequences is a sequence of conditionals which adheres to the following pattern:

(1) *Sobel sequence schematic*
$\phi \Box\!\!\rightarrow \chi$, but $(\phi \wedge \psi) \Box\!\!\rightarrow \neg\chi$

(2) If the USA threw its weapons into the sea tomorrow, there would be war; but if all the nuclear powers threw their weapons into the sea tomorrow, there would be peace. [11]

At first, they were put forth as an argument in favor of variably-strict conditional semantics [16, p. 106], since contemporary strict analyses assumed that the ϕ-conditionals would range over all worlds, including the contradicting $\phi \wedge \psi$-worlds. The situation reversed itself when Heim [4] noted that a reversal of Sobel sequences, called Heim sequences, leads to infelicity:

(3) *Reverse Sobel sequence / Heim sequence schematic*
$(\phi \wedge \psi) \Box\!\!\rightarrow \neg\chi$, but $\phi \Box\!\!\rightarrow \chi$

(4) ??If all the nuclear powers threw their weapons into the sea tomorrow, there would be peace; but if the USA threw its weapons into the sea tomorrow, there would be war.[4]

The infelicity of such sequences is highly unexpected by variably-strict models, as the world selection process is completely autonomous for each conditional, is entirely unaffected by outside influences, and only selects the closest antecedent-worlds. Therefore, building upon this initial observation, von Fintel [17] and Gillies [3] developed dynamic strict approaches that render Sobel sequences felicitous and all Heim sequences infelicitous. Thereafter, Heim sequences

*I would like to thank Maribel Romero, María Biezma, Peter Klecha, Kai von Fintel, Sven Lauer, and Irene Heim for providing valuable feedback on the content of this paper. I would also like to thank those who were kind enough to provide me with their felicity judgements on the examples within this paper. This research has been supported by the Research Unit 1614 "What if?" funded by the Deutsche Forschungsgemeinschaft (DFG).

were considered a major argument against variably-strict semantics up until the examination of felicitous Heim sequences by Moss [15]. See below for some such sequences:

(5) If kangaroos had no tails and they used crutches, they would not topple over. But if kangaroos had no tails, they would topple over.(adapted from [11, p. 1,9] by [14, p. 7])

(6) *(Holding up a dry match, with no water around)* If I had struck this match and it had been soaked, it would not have lit. But if I had struck this match, it would have lit.
(adapted from [16, p. 106] by [14, p. 7])

(7) *(Said to someone who had just been completely alone by a frozen lake)* If you had walked on the thin ice while being supported by someone on the shore, the ice wouldn't have broken. But, of course, if you had walked on the thin ice, the ice would have broken.
(adapted from [2, p. 166] by [14, p. 8])

As the dynamic strict approaches were specifically designed to enforce the infelicity of Heim sequences, such models naturally had problems accounting for their felicitous counterexamples. To respond to such findings, more and more researchers returned to variably-strict analyses of conditionals [15, 6, 14], as they initially predict all Heim sequences to be felicitous. Then, in order to account for the well-known infelicitous cases, they introduce extramodular semantic and pragmatic tools that attempt to systematically disqualify these sequences. Some such possible tools are Moss' epistemic irresponsibility [15], Klecha's modal subordination [6], need for contrastive stress [6], imprecision and precisification [6], and Karen Lewis' reordering of the world ordering according to the perceived relevance of the respective worlds [13, 14].

Another crucial observation was made by Klecha [5, 6], who argues that Sobel sequences are too vaguely defined: There are two distinct subtypes of Sobel sequences, True Sobel sequences and Lewis sequences, which may share surface similarities, but constitute two entirely independent phenomena. Their conglomeration muddled the analysis of Sobel and Heim sequences and is largely responsible for the controversially debated status of Heim sequences: True Sobel sequences are generally reversible, whereas Lewis sequences are not.

The goal of this paper is threefold. Firstly, it aims to show that neither modal subordination nor the need for contrastive stress is enough to correctly predict the infelicity of some reverse (True) Sobel sequences. Secondly, it aims to show that Klecha's imprecision-based prediction that all Lewis sequences are irreversible is too strong in light of newly acquired data that points to the contrary. Thirdly, it aims to show that Lewis' relevancy-based framework for conditionals (i) is able to naturally derive the distinction between True Sobel sequences and Lewis sequences, if we accept Bennett [2] and Arregui's [1] view on world similarity, (ii) provides a desirably weaker prediction concerning the irreversibility of Lewis sequences, and (iii) reunifies True Sobel sequences and Lewis sequences under a single semantic-pragmatic analysis.

2 True Sobel sequences, Lewis sequences, and Imprecision

Klecha [6] argues that the label *Sobel sequence* is too vague and that two similar but distinct phenomena are thereby falsely grouped together. He argues that Sobel sequences should be separated into two distinct classes: *True Sobel sequences* and *Lewis sequences*. The derivation of (in-)felicity for (reverse) True Sobel sequences is entirely independent and different from the derivation of (in-)felicity for (reverse) Lewis sequences. The difference between the two sequence types lies in the causal relation between the antecedental propositions.

(8) *True Sobel sequences*
True Sobel sequences are sequences that adhere to the following pattern: $\phi \Box\!\!\rightarrow \chi$, but $(\phi \wedge \psi) \Box\!\!\rightarrow \neg\chi$, where ϕ, ψ are causally unrelated propositions.

(9) *Lewis sequences*
Lewis sequences are sequences that adhere to the following pattern: $\phi \Box\!\!\rightarrow \chi$, but $(\phi \wedge \psi) \Box\!\!\rightarrow \neg\chi$, where ϕ, ψ are related such that ϕ precedes ψ in a causal chain of events.

Klecha [6] argues that all further differences arise from this single difference in causality relation, so long as we accept Bennett and Arregui's [2, 1] view on how causality affects world similarity: In a simplified version, they posit that the closeness of two worlds to one another is their similarity in all matters except those which pertain to the antecedent and except what follows causally from the antecedent. As such, ψ-worlds would only be counted as distant to ϕ-worlds, if ψ was not part of some causal chain that had been started by ϕ [2]. Therefore, the $\phi \wedge \psi$-worlds would count as just as close to the evaluation world w_0 as the ϕ-worlds, if ψ occurred due to a causal chain begun by ϕ [5]. The most important consequence that follows from this is Klecha's prediction that only Lewis sequences are truly irreversible (see § 2.2)

2.1 (In-)Felicity of (Reverse) True Sobel Sequences

Let us look at the True Sobel sequences, their reversals, and their respective semantics. The adoption of Bennett's [2] view on world similarity had no impact whatsoever on the way True Sobel sequences' semantics functions, when compared to the original class of Sobel sequences: In fact, Klecha posits that True Sobel sequences simply follow the conservative variably-strict models that were put forth by Stalnaker [16], Lewis [11], or Kratzer [8]. Therefore, from a strictly semantic point of view, the order of conditionals should be irrelevant, as the world selection operates on a conditional-to-conditional basis with no room for outside influences:

(10) For all contexts c, $\phi \Box\!\!\rightarrow \psi$ is true at w in c iff all the closest ϕ-worlds to w are ψ-worlds, where closeness is determined by similarity.

As such, the verse sequences (5)-(7) are correctly predicted to be felicitous. Still, this analysis presents a problem for infelicitous sequences such as (4). Here, some further mechanism is required to selectively exclude some but not all reverse True Sobel sequences [6]. Klecha argues in favor of two such possibilities: modal subordination and the need for contrastive stress.

If the second conditional of a sequence is modally subordinate to the first conditional, then the ϕ-conditional in (4) would be interpreted as *if the USA and all the nuclear powers threw their weapons into the sea, there would be war.* This interpretation would be a direct contradiction to the prior $\phi \wedge \psi$-conditional, explaining a general feeling of infelicity. However, there is no obvious reason for as to why only some conditionals are subject to modal subordination (e.g. (4)), but others are not (e.g. (5)-(7)). Another problem is the issue raised by Lewis: The felicity judgments for the same Heim sequence vary from person to person [13, 14]. To the best of our knowledge, current research does not show why modal subordination should be subject to such heavy fluctuation. As such, before these issues are addressed, this avenue is not sufficient to adequately explain the distribution of felicity for reverse True Sobel sequences.

The second possible excluding factor, the need for contrastive stress, would predict that any sequence of conditionals is infelicitous where the second antecedent has no element that is contrastively stressable against the previous antecedent. Unreversed sequences automatically have some element that can be stressed in their second conditional, since they introduce the possibility of ψ in its antecedent. This is not necessarily the case for reversed sequences:

(11) Ida: If you had stood there wearing a helmet, you wouldn't have been killed.
Aaron: # But if you had stood there, you would have been killed. [6, p.5]

Here, the antecedent of the ϕ-conditional is a syntactic subset of the $\phi \wedge \psi$-conditional's antecedent. Therefore, there is no item that could possibly be contrastively stressed in the ϕ-conditional, leading to a prediction that the reverse sequence should be infelicitous. This prediction is borne out [6]. More generalized, this approach makes two predictions: (i) Contrastively stressable sequences are felicitous barring other factors, and (ii) contrastively unstressable sequences are generally infelicitous. However, further down the line, these predictions break down rather quickly, if more data is considered: Not only are the ϕ-antecedents in (5) and (6) syntactic subsets to their respectively preceding $\phi \wedge \psi$-antecedents, without rendering the reverse sequences infelicitous, but the sequence in (4) even has an element that can be contrastively stressed (*the USA* is stressable against *all nuclear powers*), yet that sequence is considered infelicitous. As such, counterexamples to either prediction exist: (i) There are contrastively stressable sequences that are infelicitous, and (ii) there are contrastively unstressable sequences that are felicitous. We therefore rule out this approach as a viable candidate.

This would leave us, as of yet, with the correct prediction that some True Sobel sequences are reversible, but without any mechanism to correctly rule out their infelicitous counterparts.

2.2 (In-)Felicity of (Reverse) Lewis Sequences

The semantics and pragmatics of Lewis sequences, on the other hand, are quite different from the semantics of classic Sobel sequences: Since ϕ- and $\phi \wedge \psi$-conditionals would range over the same set of worlds [2], the ϕ-conditional would always be considered false, if the $\phi \wedge \psi$-conditional was considered true, regardless of the sequence order. As such, some tool needs to be introduced to allow us to ignore the $\phi \wedge \psi$-worlds for the ϕ-conditional of a non-reversed Lewis sequence (yet disallow us to ignore them for reversed ones). The tool advocated by Klecha [5, 6] is *imprecision* and *precisification*. Imprecision refers to the fact that a strictly false statement can be felicitously uttered [11], so long as the statement in question is considered "true enough" for present purposes [9]. See the example context and utterance below.

(12) *Mary arrived at work at 15:03*
Ida: Mary arrived at three o'clock. (original due Lasersohn [9])

Precisifaction refers to the act of raising the previously introduced lower standard of precision of an utterance. Once a higher standard of precision has been introduced, the lower level of precision is no longer easily accessible: Therefore, the reutterance of an imprecise statement is considered infelicitous, if precisification occurred, since precisification is generally unidirectional [12, 9, 7]. See below for an extended example of (12) that undergoes precisification.

(13) *Mary arrived at work at 15:03. This is known to all discourse participants.*
John: Mary arrived at three o'clock.
Jane: No, she arrived at 15:03.
John: # She arrived at three o'clock.

Klecha argues that Lewis sequences are handled analogously to the cases in (12) and (13). Imprecision makes their ϕ-conditionals felicitous by rendering them "true enough" via omitting the $\phi \wedge \psi$-worlds from the evaluation of the conditional. The $\phi \wedge \psi$-conditionals, on the other hand, introduce a higher standard of precision concerning the domain of worlds that is quan-

tified over. Reverse Lewis sequences are thereby rendered infelicitous, as their ϕ-conditional is uttered after a higher level of precision has already been introduced: Since precisification is unidirectional, the ϕ-conditional can no longer ignore the presence of the $\phi \wedge \psi$-worlds in its domain, rendering it contradictory to the preceding $\phi \wedge \psi$-conditional:

(14) *Construction workers Daryl, Aaron, and Ida, stand around a construction site. Daryl is not wearing a helmet. A large beam falls from above them and lands where no one was standing, but near to Daryl.*

a. Aaron: Daryl, if you had been standing there, you would have been killed.
b. Ida: But if he had been standing there and he saw the shadow of the falling beam and managed to jump out of the way in time, he would not have.
c. Aaron: # Exactly. But what I said is still right: If you had been standing there, you would have been killed. [6, p. 7]

It would therefore appear that Klecha's prediction concerning the irreversibility of Lewis sequences is accurate for the most part: At the very least, there are far fewer felicitous reverse Lewis sequences than there are felicitous reverse True Sobel sequences. Still, the new data in (15) suggests that some Lewis sequences are reversible, contrary to Klecha's prediction.

(15) *Construction workers Daryl, Aaron, and Ida, stand around a construction site. Daryl is not wearing a helmet. A large beam falls from above them and lands where no one was standing, but near to Daryl. Daryl is also known to possess exceptionally bad reflexes: Generally, 9/10 attempts to evade anything as fast as the falling beam result in failure.*

a. Aaron: Daryl, if you had been standing there, you would have been killed.
b. Ida: But if he had been standing there and he saw the shadow of the falling beam and managed to jump out of the way in time, he would not have.
c. Aaron: True, but what are the chances of THAT happening? My point stands: If he had stood there, he would have died. (adapted and modified from [6, p. 7])

There are two ways how the consistency of (15) could be reconciled with the imprecision-based framework: The first way would be to say that the interjectory probability-questioning sentence in (15-c) is effectively reversing the previous precisification. This seems like an unlikely option: Precisification is well-known to be very difficult to undo [12, 9, 10, 7]. Also, contrary to (15-c), our previous example (13) appears unable to lower the standards of precision as easily:

(16) *Mary arrived at work at 15:03. This is known to all discourse participants.*

John: Mary arrived at three o'clock.
Jane: No, she arrived at 15:03.
John: Well, okay, that's true. But who cares about those three minutes?
She arrived at three o'clock.

We therefore tentatively exclude the reversal of precisification as an explanatory candidate. The second way of how the consistency of (15) could be explained would be the conversion of the Lewis sequence into a True Sobel sequence. To do this, we would need to break the causal chain that links the two antecedental propositions together (i.e. deny that Daryl standing there could ever lead to him jumping out of the way in time). However, (15-c) does not negate the inherent possibility of Daryl's hypothetical evasive maneuver; it only questions its probability. In fact, the (improbable) possibility can be felicitously acknowledged even quite explicitly:

(17) a. Aaron: Daryl, if you had been standing there, you would have been killed.
b. Ida: But if he had been standing there and he saw the shadow of the falling beam and managed to jump out of the way in time, he would not have.
c. Aaron: Granted, but the chances of that happening are like really, really low. So my point stands: If he had stood there, he would have died.
(adapted and modified from [6, p. 7])

Therefore, we also exclude this reconciliatory possibility from being a viable candidate. In order to explain the possibility of reverse Lewis sequences such as (15) and (17), we therefore turn to a different model for conditionals: Lewis' [13, 14] relevancy-based variably-strict semantics.

3 Relevancy

Karen Lewis argues that von Fintel [17], Gillies [3], and Moss [15] were all partially right in their analysis of Sobel sequences and Heim sequences: She agrees with Moss that the effect of the first conditional on the context is pragmatic in nature, whereas she agrees with von Fintel and Gillies that this pragmatic effect has a semantic influence on the interpretation of the second conditional. That is to say, Lewis argues that infelicitous Heim sequences are not merely infelicitous, but also inconsistent. She furthermore agrees with Moss that the variably-strict Stalnaker-Lewisian framework more accurately models conditional semantics. In fact, she carries over the majority of its basic framework: The only change that is made to the traditional model is that she no longer assumes that world closeness is equated with world similarity, but rather determined by a function that incorporates both similarity and relevance. It should be noted, however, that the similarity ordering Lewis employs is Lewisian rather than Bennettian.[1] Compare the original definition in (10) with Lewis' new definition in (18):

(18) For all contexts c, $\phi \Box\!\!\rightarrow \psi$ is true at w in c iff all the closest ϕ-worlds to w are ψ-worlds, where closeness is a function of both similarity and relevance. [14, p. 20]

The essential idea behind how relevancy affects the closeness of worlds is that similarity provides the basic layout of worlds, which is then manipulated by relevancy: Low relevancy pulls worlds further away from the evaluation world, whereas high relevancy pushes less similar worlds closer to it, so that these less similar worlds are — if they are similar enough to the others — amongst the closest worlds. The relevancy of worlds, in turn, is largely manipulated by conversational context and discourse. That means that the world ordering is actively, but limitedly, determined by discourse participants: "They can indirectly affect what is (ir)relevant by changing the conversational purposes, by, for example, raising the standards of precision, making something salient, raising a new question under discussion, or refusing to accommodate a shift in conversational purpose." [14, p. 20] Of these possibilities, the raising to salience is of special import to Heim sequences. Since discourse participants must take the antecedent of a conditional seriously, in order to evaluate the counterfactual, the possibility of the antecedent is thereby automatically raised to salience [14]. This saliency can, given the right conditions, raise the relevance of the antecedent worlds. In terms of infelicitous Heim sequences, such as the one in (14), this equates to the $\phi \wedge \psi$-worlds being pushed towards the evaluation world such that the $\phi \wedge \psi$-worlds are counted amongst the closest ϕ-worlds. This general pattern for infelicitous Heim sequences is visually represented in figure 1. Since the $\phi \wedge \psi$-worlds are now

[1] Her world ordering is actually closer, by way of description, to Bennett's than it is to Lewis'. However, she herself admitted that she ignored their differences, which were not relevant to her present purposes [14, p. 8].

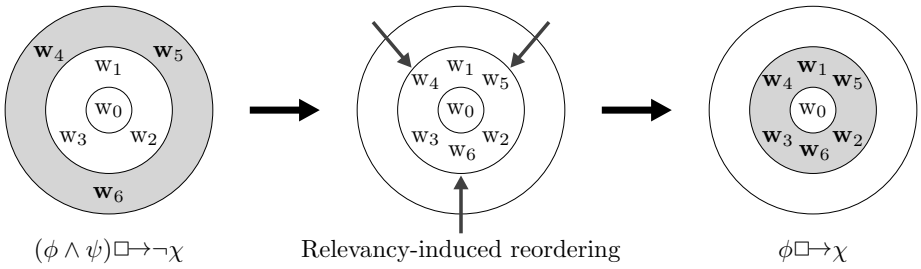


Figure 1: World ordering and selection of inconsistent Heim sequences according to Lewis [14]

just as close to the evaluation world as the ϕ-worlds, the $\phi\Box\!\!\rightarrow\chi$ conditional would also quantify over these worlds, which leads to a contradictory statement.

Not every salient world is a relevant world, however [14]: Worlds that are too dissimilar to the actual world, for example, are not raised to enough relevance, regardless of salience. In (7), for example, it was specified that the person in question was very much alone by the frozen lake. When talking about whether or not that person would have broken through the ice, had they walked upon it, the possibility of a person spontaneously appearing as if out of thin air is simply not relevant. Whilst the corresponding $\phi \wedge \psi$-worlds are certainly raised to salience, they are not relevant enough to justify pushing them to the closest ϕ-worlds. As such, no relevancy-induced restructuring of the world ordering takes place in (7), or in any of the other felicitous Heim sequences. Therefore, the ϕ-conditional does not quantify over $\phi \wedge \psi$-worlds, leading to a consistent sequence of conditionals. This is visually represented in figure 2.

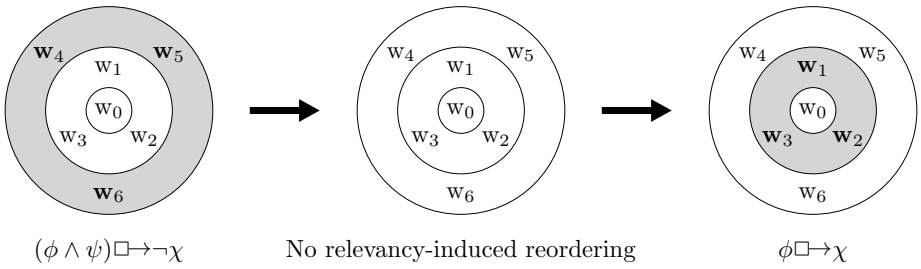


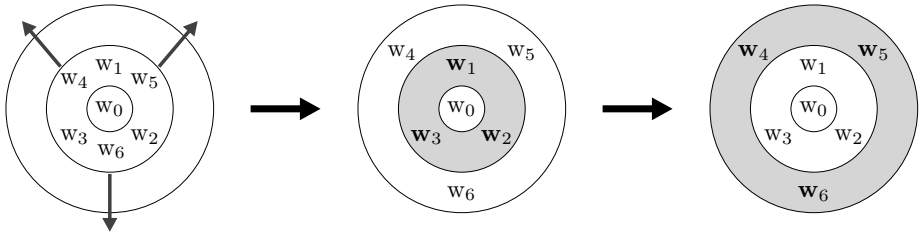
Figure 2: World orderings and selections of consistent Heim sequences according to Lewis [14]

The instability concerning the felicity judgments of Heim sequences is also predicted by this account, as its sensitivity to discourse relevancy grants the discourse participants some leeway in their semantic evaluation of the conditionals: "Hearing things at one moment as felicitous (consistent) and the next as infelicitous (inconsistent), or vice versa, is an expected feature of a phenomenon involving context sensitivity." [14, p. 22]

3.1 Klecha-Lewisian Framework

In general, we agree with Lewis [13, 14] on the nature of conditionals and the validity of her framework. However, we strongly object to the absence of a distinction between True Sobel sequences and Lewis sequences within her framework: Whilst Klecha's [5, 6] predictions were arguably too strong, as seen in § 2.2, his observation that reverse Lewis sequences are far more likely to be infelicitous holds true. We therefore need some way to incorporate parts of

Klecha's analysis into Lewis' contextualist framework. To do this, we start with the same basic assumption that was required by Klecha's analysis: Bennett's [2] and Arregui's [1] view on world ordering. The incorporation of their work into Lewis' framework has only one currently relevant impact: The similarity ordering of Lewis sequences is such that ϕ-worlds and $\phi \wedge \psi$-worlds are equally similar to the evaluation world. Contrary to Klecha's model, this poses no immediate issue, since similarity is no longer the sole determining factor for world closeness. Assuming that low relevancy pulls these $\phi \wedge \psi$ worlds further away from the evaluation world, these worlds would no longer be counted amongst the closest ϕ-worlds for the evaluation of the ϕ-conditional in a Lewis sequence. This assumption of low relevance appears very intuitive: Moss [15], Klecha [6, 5], and Lewis [14] all make the same assumption in one way or another (implicitly or explicitly). Klecha, in particular, requires the implicit assumption that $\phi \wedge \psi$-worlds are contextually less relevant than the ϕ-worlds to motivate the low level of precision a Lewis sequence starts out with.[2] Lewis, on the other hand, explicitly states that certain possibilities can be considered contextually irrelevant for discourse purposes (i.e. the speaker trying to make a point) until some discourse participants brings them into play [14, p. 21]. Whilst she was talking about Sobel sequences in general, it certainly fits the description of what appears to be happening to Lewis sequences. Once these worlds are pulled further away from the evaluation world by their contextual irrelevancy, the remaining evaluation of the sequence is true to the standard variably-strict analysis, as is seen in figure 3. In this, Lewis sequences differ



Relevancy-induced reordering $\phi \Box\!\!\rightarrow \chi$ $(\phi \wedge \psi) \Box\!\!\rightarrow \neg\chi$

Figure 3: Proposed world closeness orderings and world selections of Lewis sequences

from the analysis of True Sobel sequences: In order to make the ϕ-conditional a true statement, Lewis sequences require low relevancy to interfere with the similarity ordering, whereas True Sobel sequences require nothing of the sort.

Having demonstrated that Lewis sequences pose no immediate problem, we turn to their reverse counterparts. There are two ways a reverse Lewis sequence can be judged as infelicitous: A pure reverse Lewis sequence requires no special steps. The initial discourse context acknowledges the relevancy of the $\phi \wedge \psi$-worlds and thereby does not pull them further away from the evaluation world. The ϕ-conditional also ranges over the $\phi \wedge \psi$-worlds, leading to a contradictory statement. A more interesting case is the reverse Lewis sequence in (14), where the reverse Lewis sequence is embedded within a standard Lewis sequence. Since the $\phi \wedge \psi$-worlds were originally moved further away from the evaluation world, they need to be pulled back in, in order to make the reverse Lewis sequence inconsistent. Whether or not the $\phi \wedge \psi$-worlds are counted amongst the closest ϕ-worlds is then dependent on the same criteria that Lewis [14] originally posited: (i) Their possibility needs to be salient, (ii) they must be similar enough to the other closest ϕ-worlds, and (iii) they must be counted as relevant for the purposes of the discourse. The first criteria is automatically fulfilled, as the possibility of an antecedent is

[2] Lewis actually states that low precision is comparable to lower relevancy in her framework [14, p. 20].

always raised to salience. The second criteria is also automatically fulfilled, since $\phi \wedge \psi$ worlds and ϕ-worlds are equally similar in Lewis sequences. Therefore, the sole deciding factor for Lewis sequences is the relevancy to the current discourse. This criteria is also, in most cases, automatically fulfilled: We would argue that any question under discussion that considers it relevant whether or not χ would follow from ϕ would also be sensitive to any possibility ψ that is directly or indirectly caused by ϕ and that could possibly prevent χ. Positing all of Lewis' criteria would also predict, however, that the worlds in question must not be intrinsically irrelevant: They must be considered at least realistic, even if highly improbable, by the discourse participants. We would therefore predict that some reverse Lewis sequences are consistent, even if no explicit questioning of the relevance of $\phi \wedge \psi$ worlds takes place (as was indirectly done in (17-c)). This prediction appears to be borne out, considering the reverse Lewis sequence below:

(19) a. A: If I had dropped that vase, it would have broken.
b. B: But if you had dropped that vase and that drop caused it to quantum-tunnel to a cushy pillow, it would not have.
c. A: Okay, but what I said is still true: If I had dropped it, it would have broken.

Which leads us to our other original examples in (15) and (17). The explanation of their felicity itself is simplistically straightforward within this framework. Both sequences either question the probability of the $\phi \wedge \psi$-worlds or explicitly asserted their improbable nature. In most cases, probability and relevance are almost intrinsically tied together. By questioning their probability, the speaker also questioned their relevancy to the discourse. In doing so, the discourse participant pushes the $\phi \wedge \psi$-worlds further away from the evaluation world, again, which leaves them free to reassert their original conditional without inconsistency. See below:

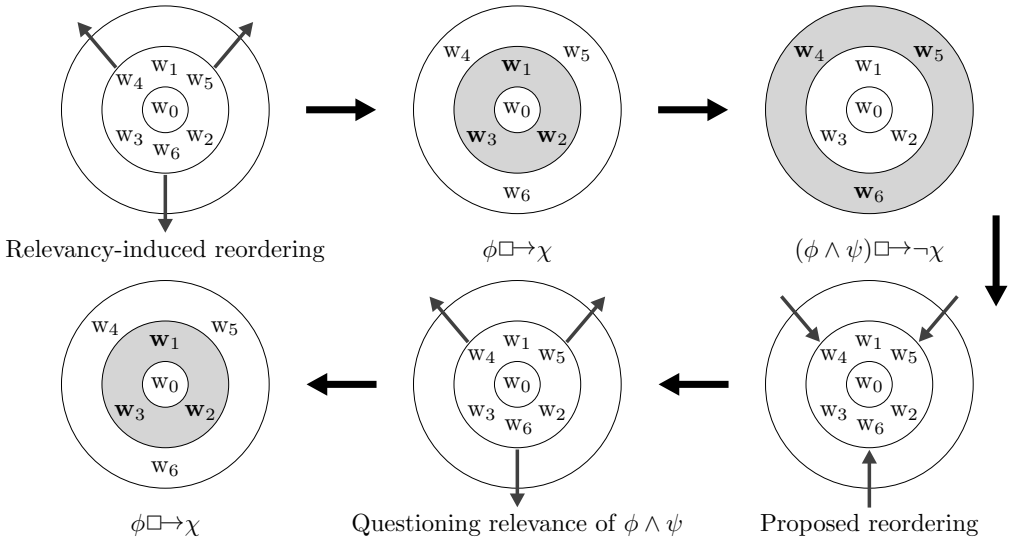


Figure 4: Proposed world closeness orderings and world selections of (15) and (17)

4 Conclusion

In summary, we have shown that Klecha's [5, 6] predictions concerning the unidirectionality of Lewis sequences is too strong, that Lewis sequences are reversible if the probability or relevance of their $\phi \wedge \psi$-worlds is questioned, and that Lewis sequences are reversible if their $\phi \wedge \psi$-worlds are intrinisically irrelevant. We have also argued that Klecha's account for the analysis of infelicitous reverse True Sobel sequences is unsatisfactory, due to its lack of (correct) predictive power. On the other hand, we have argued that Lewis' [13, 14] framework handles Sobel sequences well, but lacks the necessary distinction between True Sobel sequences and Lewis sequences. In adopting Bennett's [2] and Arregui's [1] view on the impact of causality on world similarity, we have shown that a Klechaesque analysis of Lewis sequences follows naturally, albeit with the desired weaker prediction of unidirectionality. The fact that reverse Lewis sequences are usually infelicitous is explained by their automatic fulfillment of two thirds of the necessary criteria, plus the near-guaranteed fulfillment of the only remaining third criteria (the relevance of $\phi \wedge \psi$ to the current question under discussion). Not only that, but our framework treats Lewis sequences and True Sobel sequences as proper subsets of a single phenomenon (Sobel sequences), rather than as two entirely separate phenomena with coincidental surface similarities which require two entirely different explanations [5, 6].

For future research, we believe it absolutely necessary and vital to finally experimentally verify or falsify the predictions that the different models have made about the felicity and consistency of reverse True Sobel sequences and reverse Lewis sequences over the years.

References

[1] Ana Arregui. On similarity in counterfactuals. *Linguistics and Philosophy*, 32(3):245–278, 2009.

[2] Jonathan Bennett. *A Philosophical Guide to Conditionals*. Oxford University Press, Oxford, 2003.

[3] Anthony S. Gillies. Counterfactual scorekeeping. *Linguistics and Philosophy*, 30(3):329–360, 2007.

[4] Irene Heim. Unpublished MIT seminar on conditionals, 1994.

[5] Peter Klecha. *Bridging the Divide: Scalarity and Modality*. PhD thesis, University of Chicago, 2014.

[6] Peter Klecha. Two Kinds of Sobel Sequences: Precision in Conditionals. In U. Steindl, T. Borer, H. Fang, A. G. Pardo, P. Guekguezian, B. Hsu, C. O'hara, and I. C. Ouyyang, editors, *Proceedings of the West Coast Conference on Formal Linguistics 32*, 2014.

[7] Peter Klecha. On unidirectionality in precisification. *Linguistics and Philosophy*, To appear.

[8] Angelika Kratzer. Modality. In A. von Stechow and D. Wunderlich, editors, *Semantics: An international handbook of contemporary research*. De Gruyter, 1991.

[9] Peter Lasersohn. Pragmatic Halos. *Language*, 75:522–551, 1999.

[10] Sven Lauer. On the pragmatics of pragmatic slack. 2:389–401, 2012.

[11] David K. Lewis. *Counterfactuals*. Blackwell, Malden, 1973.

[12] David K. Lewis. Scorekeeping in a language game. *Journal of Philosophical Logic*, 8(1):339–359, 1979.

[13] Karen Lewis. Elusive Counterfactuals. *Noûs*, 50(2):286–313, 2016.

[14] Karen Lewis. Counterfactual Discourse in Context. *Nous*, To appear.

[15] Sarah Moss. On the pragmatics of counterfactuals. *Noûs*, 46(3):561–586, 2012.

[16] Robert C. Stalnaker. A Theory of Conditionals. In N. Rescher, editor, *Studies in Logical Theory*, pages 98–112. Basil Blackwell Publishers, Oxford, 1968.

[17] Kai von Fintel. Counterfactuals in a Dynamic Context. In M. Kenstowicz, editor, *Ken Hale: A Life in Language*, pages 123–152. MIT Press, Cambridge, 2001.

'I believe' in a ranking-theoretic analysis of 'believe'*

Sven Lauer

University of Konstanz, Konstanz, Germany
sven.lauer@uni-konstanz.de

Abstract

There is a *prima facie* tension between two well-known observations about sentences of the form *I (don't) believe p*. The first is *Moore's paradox*, i.e., the fact that sentences of the form *p, but I don't believe p* and $\neg p$ *but I believe p* sound 'contradictory' or 'incoherent'. The second observation is that *I believe* often functions as a hedge: A speaker who asserts *I believe p* often (but not always) conveys that she is not certain that *p*, or that she does not want to commit entirely to *p* being true.

I argue that (natural explantations of) these two observations are in conflict with respect to the following question: Does saying *I believe p* commit the speaker to taking *p* to be true? Moore's paradox says *Yes!*, while the fact that *I believe* functions as a hedge says *No!* I argue that resolving this tension (along with other desiderata) requires a theory of *graded belief*. I reject a probabilistic threshold analysis for familiar reasons (lack of closure under conjunction), and show that an alternative based on the *ranking theory* of Spohn (1988, 1990, 2012) can compositionally deliver the desired results when embedded in a commitment-based theory of declarative force.

1 Two observations and a dilemma

1.1 Moore's paradox

Moore's paradox is the observation that (1) and (2) sound 'incoherent' or 'contradictory'.

(1) It is raining, but I don't believe it (is raining). $p \wedge \neg \mathsf{Bel}_{\mathrm{Sp}}(p)$ or $p \wedge \mathsf{Bel}_{\mathrm{Sp}}(\neg p)$

(2) It is not raining, but I believe it (is raining). $\neg p \wedge \mathsf{Bel}_{\mathrm{Sp}}(p)$

(1) has two readings, depending on the relative scope of negation and *believe*. For simplicity, in the following, I will talk mostly about (2), which is equally Moore-paradoxical. The account offered in the second half of this paper, however, also extends to the two readings of (1).

These sentences sound contradictory, even though they should, given standard assumptions about their meanings, express perfectly consistent propositions. One way to see is this to consider the sentences in (3).

(3) a. It is not raining, but John believes/thinks it is (raining).
b. It was not raining, but I believed/thought it was (raining).

These sentences are consistent because it is possible that John has false beliefs, and it is possible that the speaker used to have false beliefs. But, of course, it is also possible that the speaker presently has false beliefs, so (2) should have a consistent content, as well. And yet, (2) sounds contradictory.

Here is a sketch of a natural explanation of these observations: With uttering $\neg p$, a speaker commits to taking *p* to be false, but with uttering *I believe p*, she commits to taking *p* to be true. Thus, while (2) has a consistent content, it gives rise to incompatible commitments for the speaker. Hence, it sounds contradictory.

*I'd like to thank Cleo Condoravdi, Ciyang Qing, Eric Raidl, Wolfgang Spohn, as well as audiences at the Konstanz *What if?*-group and the *Questioning Speech Acts* workshop, for helpful comments and discussion.

1.2 Hedging with 'I believe'

When a speaker utters *I believe p*, she often conveys that she is not entirely sure that p is true.

(4) I believe it is raining. $\rightsquigarrow$ Speaker is not sure that it is raining.

However, this is by no means always the case. The intuitive implication in (4) can be coherently denied (5a), it can be suspended by inserting an adverb like *firmly* (5b) and even without such an indication, the implication of uncertainty can be absent if the context is right.

(5) a. I believe that the president should be impeached. I am positive/absolutely certain.
b. I firmly believe that the president should be impeached.

A natural explanation of these facts is the following: The inference in (4) is a *conversational implicature*, which roughly is derived at follows. The speaker could have asserted p instead of uttering *I believe p*, which is longer and more complex. She must have had a reason to opt for *I believe p*. A plausible reason (in the right context) is that she did not want to commit to p, because she is not sure that p is true. Hence she opted for only committing to *I believe p* instead of p.

1.3 A dilemma

The two 'natural explanations' just sketched are very intuitive individually, but they are problematic in so far as they appear to presuppose opposite answers to the question 'Does uttering *I believe p* commit the speaker to taking p to be true?'

To bring this out more clearly, let $\mathsf{Ass}_{\mathrm{Sp}}(\cdot)$ be an operator that represents the normative consequences of assertion ('doxastic commitment' / 'commitment to believe' Condoravdi and Lauer 2011, Lauer 2013; 'assertoric commitment' Krifka 2014; 'truth commitment' Searle 1969, Krifka 2015; ...) and let $\mathsf{Bel}_{\mathrm{Sp}}(\cdot)$ be an operator representing the content of *I believe* statements. The question before us is whether the two operators should support the following principle ($\rightarrow$ is material implication):[1]

(6) **Mixed extraspection**: $\mathsf{Ass}_{\mathrm{Sp}}(\mathsf{Bel}_{\mathrm{Sp}}(\phi)) \rightarrow \mathsf{Ass}_{\mathrm{Sp}}(\phi)$

Note that (6) is independent, in principle, from both (7) and (8):

(7) **Extraspection for belief**: $\mathsf{Bel}_{\mathrm{Sp}}(\mathsf{Bel}_{\mathrm{Sp}}(\phi)) \rightarrow \mathsf{Bel}_{\mathrm{Sp}}(\phi)$

(8) **Extraspection for commitment**: $\mathsf{Ass}_{\mathrm{Sp}}(\mathsf{Ass}_{\mathrm{Sp}}(\phi)) \rightarrow \mathsf{Ass}_{\mathrm{Sp}}(\phi)$

(7) and (8) are intuitively plausible. (7) says that an agent cannot mistakenly think he has a belief. In a Kripke-model for $\mathsf{Bel}_{\mathrm{Sp}}(\cdot)$, (7) corresponds to density of the accessibility relation ($\forall w_1, w_2 : w_1 R w_2 \rightarrow \exists v : w_1 R v \wedge v R w_2$), which follows from Euclideanity ($\forall w, v_1, v_2 : wRv_1 \wedge wRv_2 \rightarrow v_1 R v_2$). Euclideanity, in turn, correponds to the following principle, which is standardly assumed for belief:

(9) **Negative introspection**: $\neg\mathsf{Bel}_{\mathrm{Sp}}(\phi) \rightarrow \mathsf{Bel}_{\mathrm{Sp}}(\neg\mathsf{Bel}_{\mathrm{Sp}}(\phi))$

(8) says that an agent who is committed to being committed to p is automatically committed to p. This seems also very plausible, and the principle plays a crucial role in Condoravdi and Lauer's (2011) analysis of explicit performatives.

[1] I borrow the term 'extraspection' from van der Hoek (1993). Rieger (2015) calls the principle 'positive belief infallibility'.

But what about the mixed extraspection principle in (6)? Here we are in a bind. The 'natural explanation' for the fact that *I believe* is used as a hedge presupposes that (6) is *not valid* (else, saying *I believe p* is not a way of avoiding commitment to *p*). On the other hand, the 'natural explanation' for Moore's paradox apparently presupposes that (6) *is valid* (else, Moore-paradoxical sentences do not give rise to contradicting commitments).

In the rest of this paper, I will work towards a compositional analysis of *believe*-sentences that avoids this dilemma while maintaining the intuitive core of the 'natural explanations' sketched above. In the present context, the desideratum of compositionality amounts to the following two requirements: First, the sentences in (10) should get the same kind of content, modulo the belief subject and tense.

(10) a. I believe it is raining.
b. John believes it is raining.
c. I believed that it was raining.

Second, *'p'* and *'I believe p'* should be assigned a uniform (declarative) force, with the different implications of the two kinds of sentences tracing to their (different) contents.

2 Diagnosis: Four desiderata

Intuitively, if we want to account for hedging-effects as an implicature, then **mixed extraspection** (6) must fail because, in the titular slogan of Hawthorne et al. (2015), 'belief is weak'. Their slogan (and central thesis) gives rise to a first desideratum for a theory of belief (self-)ascriptions:

(11) **Weakness.** Assertion of '*I believe p*' induces a weaker commitment than assertion of '*p*'.

At the same time, however, the commitment induced by belief self-ascriptions should not be too weak. It must be strong enough to deliver on the following two desiderata:

(12) **Moore's paradox.**
Assertion of '*p, but I don't believe p*' gives rise to inconsistent commitments.

(13) **Consistency.**
Assertions of '*I believe p*' and '*I believe* $\neg p$' jointly give rise to inconsistent commitments.

Consistency in particular requires that the commitment induced by *I believe p* must be stronger than the one induced by *might p*, as witnessed by the non-contradictoriness of (14).

(14) It might be raining, and/but it might also not be raining.

So we want *I believe p* to induce a stronger commitment than *might p*, but a weaker commitment than *p*. This motivates employing a representation of belief that allows for more than two grades of belief. A popular theory of this kind is Bayesian probability. Setting aside the issue of compositionality for a moment, suppose that assertoric commitments are represented as constraints on the speaker's subjective probability distribution P_{Sp} and suppose that assertion of '*p*' and 'I believe *p*' induce the following commitments, for some $\theta_{\mathsf{a}}, \theta_{\mathsf{b}} > 0.5$ (cf. Swanson 2006, Lassiter 2017 on epistemic *must*):

(15) '*p*' induces the commitment $P_{\mathrm{Sp}}(p) \geq \theta_{\mathsf{a}}$.
(16) 'I believe *p*' induces the commitment $P_{\mathrm{Sp}}(p) \geq \theta_{\mathsf{b}}$.

Such a theory meets the desiderata **Moore's paradox** (12) and **consistency** (13), since P_{Sp} cannot assign probability > 0.5 to both p and $\neg p$. However, as it stands, a probabilistic threshold analysis can account for at most one of **weakness** (11) and the following desideratum:[2]

(17) **Closure.**
Assertions of '*I believe* p' and '*I believe* q' commit the speaker to '*I believe* $p \wedge q$'.

To account for **weakness**, it must be that $\theta_{\mathsf{b}} < \theta_{\mathsf{a}} \leq 1$. But then $\theta_{\mathsf{b}} < 1$, and hence **closure** is unaccounted for, because it is always possible to assign probabilities to p and q such that they are larger than θ_{b}, but their conjunction has a probability smaller than θ_{b}.[3]

3 Ranking Theory

We want a theory of graded belief that, together with a theory of assertion, meets all four desiderata: **weakness** (11), **Moore's paradox** (12), **consistency** (13) and **closure** (17).

In the following, I will spell out such a theory in a compositional fashion, employing the *ranking theory* of Spohn (1988, 1990, 2012). The basic construct for the representation of beliefs in the version of ranking theory I am going to use here is that of a *ranking mass function*.[4]

Definition 1 (Ranking mass function, after Spohn 2012, p. 70)**.** *Given a set of worlds W, a* ranking mass function *is any function $k : W \to (\mathbb{N} \cup \{\infty\})$ such that $k^{-1}(0) \neq \emptyset$.*

The set of all ranking mass functions over W is denoted by $\mathbb{K}_W$.

Raidl (t.a.) articulates well how ranking mass functions should be interpreted:

> "One may think of a ranking mass [function] as a doxastic ordering source. The zero worlds are the closest worlds, or the best candidates for the actual world. The greater the rank of a world, the less plausible that world is as a candidate for the actual world, the more it is disbelieved or the more the agent has doubts about it. Worlds ranked with $n < \infty$ are within the doxastic modal horizon (or modal base). Worlds with rank ∞ are crazy worlds outside the modal horizon. Although the agent acknowledges their eventual possibility, she disregards them for matters of actual judgements." (Raidl t.a.)

Even though all our definitions could be stated in terms of ranking mass functions, it is convenient to define the derived notion of a *ranked belief function*.

Definition 2 (Ranked belief functions, after Spohn 2012, p. 75)**.** *Given a ranking mass function k, the* corresponding ranked belief function *β_k is that function $\wp(W) \to (\mathbb{N} \cup \{\infty\})$ such that*

[2] Whether rational belief must satisfy a constraint analogous to **closure** is of some debate in philosophical epistemology, most notably in discussions of the 'lottery paradox' (Kyburg 1961) and 'preface paradox' (Makinson 1965). For a recent discussion of these issues, and a vote for **closure** see Leitgeb (2014).

Here I confine myself to noting that, as an observation about sincere assertion (rather than rational belief), (17) appears to me unassailable: Someone who sincerely asserts *I believe p* and *I believe q*, but then (immediately) goes on to deny or withhold assent from *I believe* $p \wedge q$ clearly has failed up to live up to the commitments he has taken on with his two assertions.

[3] There are more articulated probabilistic analyses of belief that may well satisfy all our desiderata when combined with a suitable analysis of declarative force. Of particular relevance are the proposals in Lin and Kelly (2012) and Leitgeb (2015), which feature probabilistic proposals that can deliver **closure**.

[4] Here and throughout, the notation and some of the terminology departs from previous presentations of this work (including the submitted abstract) to harmonize with other recent presentations of ranking theory, especially Raidl (t.a., 2017). What Raidl (and I) call a 'ranking mass (function)' is called a 'complete pointwise ranking function' in Spohn's work.

(a) $\beta_k(W) = \infty$

(b) $\beta_k(\emptyset) = 0$

(c) *for all non-emtpy* $A \subseteq W$: $\beta_k(A) = \min\{k(w) \mid w \notin A\}$

A ranked belief function assigns to each proposition a measure of its belief (rather than disbelief): The higher the rank of a proposition is, the more it is believed. A crucial property of ranked belief functions is the following, which follows directly from Definition 2:

Fact 3. *For any ranked belief function* β_k *and any proposition* A*: At most one of* $\beta_k(A)$ *and* $\beta_k(W - A)$ *can be larger than* 0.

Ranked belief functions have a number of further notable properties:

Fact 4 (Properties of ranked belief functions)**.** *Any ranked belief function* β_k *is a 'positively minimizing ranking function' in the sense of Spohn (2012). That is:*

(a) $\beta_k(W) = \infty$
The tautology is always absolutely believed.

(b) $\beta_k(\emptyset) = 0$
The contradictory proposition is never believed to a positive degree.)

(c) *For all propositions* A, B: $\beta_k(A \cap B) = \min(\beta_k(A), \beta_k(B))$.
The rank of an intersection of two propositions is the smaller of the rank assigned to the individual propositions.

4 The object language: Syntax, semantics and pragmatics

For simplicity, I work with a simple propositional language, enriched with a family of modal operators for belief. This section spells out its syntax, semantics, and pragmatics.

4.1 Syntax

Definition 5 (Language)**.** *Let* P *and* I *be disjoint sets (of proposition letters and individuals, resp.). Then* $\mathcal{L}_{P,I}$ *is the smallest set such that*

1. $P \subseteq \mathcal{L}_{P,I}$ *(proposition letters)*
2. *If* $\phi \in \mathcal{L}_{P,I}$*, then* $\neg\phi \in \mathcal{L}_{P,I}$. *(negation of formulas)*
3. *If* $\phi, \psi \in \mathcal{L}_{P,I}$*, then* $(\phi \wedge \psi) \in \mathcal{L}_{P,I}$. *(conjunction of formulas)*
4. *If* $\phi \in \mathcal{L}_{P,I}$ *and* $i \in I$*, then* $(\mathsf{Bel}_i(\phi)) \in \mathcal{L}_{P,I}$. *(belief formulas)*

Other connectives are introduced as the usual abbreviations.

4.2 Semantics

Models are standard possible-worlds ones, and the interpretation of non-belief formulas is a standard one. In order to interpret the belief operator, we add two elements to the models: A function K that specifies the believe state for each agent at each world, and a threshold b.[5]

[5] In a system for English, it might be more appropriate to have b provided by the context instead of fixing it lexically. It would also likely be desirable to give the predicate *believe* a degree argument, to account for the fact that *believe* is compatible with what looks like degree modification, as in *I firmly believe p*.

Definition 6 (Models). *A **model** for $\mathcal{L}_{P,I}$ is a qadruple $M = \langle W, V, K, \mathsf{b}\rangle$, such that*

1. *W is a set of possible worlds,*
2. *$V : P \to \wp(W)$ is a valuation for the proposition letters.*
3. *$K : I \times W \to \mathbb{K}_W$ is a function that assigns to each individual-world pair a ranking mass function.*
4. *$0 < \mathsf{b} \in \mathbb{N}$ is the threshold for belief ascriptions.*

Definition 7 (Denotation function). *Given a model $M = \langle W, V, K, \mathsf{b}\rangle$, the **denotation function** $[\![\cdot]\!]^M : \mathcal{L}_{P,I} \to \wp(W)$ is as follows:*

1. *$[\![p]\!]^M = V(p)$ for all $p \in P$.*
2. *$[\![\neg\phi]\!]^M = W - [\![\phi]\!]^M$.*
3. *$[\![\phi \wedge \psi]\!]^M = [\![\phi]\!]^M \cap [\![\psi]\!]^M$.*
4. *$[\![\mathsf{Bel}_i(\phi)]\!]^M = \{w \in W \mid \beta_{K(i,w)}(\phi) \geq \mathsf{b}\}$*

In the following, I will generally omit the model parameter from $[\![\cdot]\!]$.

4.2.1 Admissibility

In order to guarantee some desirable properties, I introduce a notion of *admissibility* for ranking mass functions and models:[6]

Definition 8 (Admissibility of ranking mass functions and models). *Given a language $\mathcal{L}_{P,I}$, model $M = \langle W, V, I, \mathsf{b}\rangle$, a pointwise ranking mass function k is* admissible for $i \in I$ in M *iff* $\forall v \in k^{-1}(0) : K(i, v) = k$
A model is admissible iff $\forall w \in W, i \in I : K(i, w)$ is admissible for i in M.

Admissibility will play a crucial role in the pragmatics, defined in the next subsection. A semantic consequence of requiring admissibility is that it ensures extraspection for belief.

Fact 9 (Extraspection for belief). *For any admissible model M, for all $i \in I, \phi \in \mathcal{L}_{P,I}$:* $[\![\mathsf{Bel}_i(\mathsf{Bel}_i(\phi))]\!] \subseteq [\![\mathsf{Bel}_i(\phi)]\!]$

4.3 Pragmatics

4.3.1 Commitment frames and states

The commitments an agent has are represented as *constraints on ranking functions.*

Definition 10 (Commitment frames). *A* commitment frame *for $\mathcal{L}_{P,I}$ is any $\mathfrak{C}_\mathsf{i} = \langle M, i, \mathsf{a}\rangle$:*

1. *$M = \langle W, V, K, \mathsf{b}\rangle$ is a model for $\mathcal{L}_{P,I}$.*
2. *$i \in I$ is an individual.*
3. *$\mathsf{b} < \mathsf{a} \in \mathbb{N}$ is the* assertion threshold.

[6] For a systematic investigation of a 'ranked semantics' of the type used here, including correspondence results, see Raidl t.a., 2017).

Definition 11 (Commitment states)**.** *Given a commitment frame* $\mathfrak{C}_i = \langle M, i, \mathsf{a}\rangle$*, a* commitment state C_i *is partial truth function of ranking mass functions such that* $C_i(k)$ *is defined iff* k *is admissible for* i *in* M*. There are two distinguished commitment states:*

(a) $\bot = \lambda k.0$ *(the contradictory state)*

(b) $\top = \lambda k.1$ *(the uncommitted state)*

4.3.2 Updates for commitment states

The commitment dynamics is essentially the same as in Veltman's (1996) *Update Semantics*.[7]

Definition 12 (Declarative update)**.** *Given a commitment frame* $\mathfrak{C}_i = \langle M, i, \mathsf{a}\rangle$*, the* update function $+$ *is the following function from commitment states and formulas to commitment states:*

$$C + \phi = \lambda k.C(k) \;\&\; \beta_k([\![\phi]\!]) \geq \mathsf{a}$$

Definition 13 (Support)**.** *For any commitment state* C *and formula* ϕ*:*

$$C \vDash \phi \textit{ iff } C + \phi = C$$

We immediately obtain the following minimal requirement for a notion of commitment. If an agent is committed, in C_i to a proposition ϕ, then update with any proposition ψ that is incompatible with ϕ results in the inconsistent state.

Fact 14 (Inconsistency of commitments)**.** *Let* C_i *be a commitment state, and for some* ϕ, ψ *such that* $[\![\phi]\!] \cap [\![\psi]\!] = \emptyset$ *and* $C_i \vDash \phi$*. Then* $C_i + \psi = \bot$*.*

Proof. Follows immediately from the following Fact 15. □

Further, we can show that update with a conjunction is equivalent to subsequent update with the two conjuncts:

Fact 15 (Conjunction)**.** *For any commitment state* C_i *and formulas* ϕ, psi *:* $C_i + (\phi \wedge \psi) = C_i + \phi + \psi$*.*

Proof. First, note that

(*) For any k : $\beta_k([\![\phi \wedge \psi]\!]) = \beta_k([\![\phi]\!] \cap [\![\psi]\!]) = \min(\beta_k([\![\phi]\!]), \beta_k([\![\psi]\!]))$. (by Fact 4c)

We have to show that for any commitment state C_i and ranking mass function k: $[C_i + \phi \wedge \psi](k) = 1 \Leftrightarrow [C_i + \phi + \psi](k) = 1$. ($\Rightarrow$) Suppose that $[C_i + \phi \wedge \psi](k) = 1$, then $\beta_k([\![\phi \wedge \psi]\!]) \geq \mathsf{a}$, hence by (*) $\min(\beta_k([\![\phi]\!]), \beta_k([\![\psi]\!])) \geq \mathsf{a}$ and so $k([\![\phi]\!]) \geq \mathsf{a}$ and $k([\![\psi]\!]) \geq \mathsf{a}$. Further, since $[C_i + \phi \wedge \psi](k) = 1$, it must be that that $C_i(k) = 1$. But then $[C_i + \phi \wedge \psi](k) = 1$. ($\Leftarrow$) Suppose $[C_i + \phi + \psi](k) = 1$. Then $C_i(k) = 1$ and $k([\![\phi]\!] \geq \mathsf{a}$ and $k([\![\psi]\!]) \geq \mathsf{a}$. But then $\min(\beta_k([\![\phi]\!]), \beta_k([\![\psi]\!])) \geq \mathsf{a}$, hence by (*) $\beta_k([\![\phi \wedge \psi]\!]) \geq \mathsf{a}$. But then $[C_i + \phi \wedge \psi](k) = 1$. □

[7] I could have brought this out even more clearly by letting commitment states be sets of ranking functions. Then the contradictory state would be $\emptyset$, the uncommitted state would be $\{k \mid k \text{ is admissible for } i\}$, and the update operation $+\phi$ would be intersection with $\{k \mid \beta_k(\phi) > \mathsf{a}\}$.

For present purposes, such a set-based representation would have been sufficient. I opted for the truth-function representation instead because it generalizes better to additional phenomena (not treated here) and because it highlights the idea that commitment states are to be thought of as constraints on ranking functions.

4.4 Predictions

4.4.1 Desideratum 1: Closure

Fact 16 (Closure explained)**.** *For any commitment state C_i:* $C + (\mathsf{Bel}_i(\phi)) + (\mathsf{Bel}_i(\psi)) \models \mathsf{Bel}_a(\phi \wedge \psi)$.

Proof. Direct corollary of Fact 15. □

4.4.2 Desideratum 2: Consistency

Fact 17 (Consistency)**.** *For any commitment state C_i and any formula ϕ:* $C_i + \mathsf{Bel}_i(\phi) + (\mathsf{Bel}_i(\neg\phi)) = \bot$

Proof. Obviously, $C_i + \mathsf{Bel}_i(\phi) \models \mathsf{Bel}_i(\phi)$ and $[\![\mathsf{Bel}_i(\phi)]\!] \cap [\![\neg\mathsf{Bel}_i(\phi)]\!]$. But then, by Fact 14, $C_i + \mathsf{Bel}_i(\phi) + (\mathsf{Bel}_i(\neg\phi)) = \bot$. □

4.4.3 Desideratum 3: Weakness

Fact 18 (Weakness)**.** $C_i + \mathsf{Bel}_i(\phi) \nvDash \phi$.

Proof. Let $\mathfrak{C}_i = \langle M, i, \mathsf{a}\rangle$ with $M = \langle W, V, K, \mathsf{b}\rangle$ such that there are $w, v \in W$ such that $\beta_{K(i,w)}(V(p)) = \mathsf{b}$ and $\beta_{K(i,v)}(V(p)) = \mathsf{a}$. Recall that it must be that $\mathsf{a} > \mathsf{b}$. Then $\top + \mathsf{Bel}_i(p)$ is true of $K(i,w)$ and $K(i,v)$, but $\top + \mathsf{Bel}_i(p) + p$ is true of $K(i,v)$, but not of $K(i,w)$. But then $\top + \mathsf{Bel}_i(p) \neq \top + \mathsf{Bel}_i(p) + p$ and hence $\top + \mathsf{Bel}_i(p) \nvDash p$. □

Note that even though we require that $\mathsf{b} < \mathsf{a}$, of course nothing prevents an agent who assigns rank a to ϕ from asserting $\mathsf{Bel}_i(\phi)$. Formally: A ranking mass function k such that $\beta_k([\![\phi]\!]) \geq \mathsf{a}$ is compatible with commitment state $C_i + \mathsf{Bel}_i(\phi)$ (provided that k is compatible with the original C_i). This is as it should be: Recall that the 'natural explanation' for hedging with *I believe* that we started out treats the hedging-inference as a conversational implicature. Thus the update with $\mathsf{Bel}_i(\phi)$ should not preclude that the agent is absolutely certain that ϕ, it should only be compatible with her not being absolutely certain.

4.4.4 Desideratum 4: Moore's paradox

Fact 18 assures us that, in our system **mixed extraspection** fails. At the same time, Lemma 19 assures us that belief is not *too* weak: Declarative update with *I believe that* ϕ does not induce full assertoric commitment to ϕ (as **mixed extraspection** would have it), but it *does* induce a commitment to ϕ that goes over and above the commitment to $\mathsf{Bel}_i(\phi)$.

Lemma 19 (Belief update)**.** *Let $\mathfrak{C}_i = \langle M, i, \mathsf{a}\rangle$ with $M = \langle W, V, K, \mathsf{b}\rangle$. Then for any commitment state C_i and formula ϕ:*

$$\textit{For any } k\textit{: if } [C_i + \mathsf{Bel}_i(\phi)](k) = 1 \textit{ then } k([\![\phi]\!]) \geq \mathsf{b}$$

Proof. Suppose that for some $k : [C_i + \mathsf{Bel}_i(\phi)](k) = 1$. Then it must be that $\beta_k([\![\mathsf{Bel}_i(\phi)]\!] \geq \mathsf{a} > 0$, which implies that $k^{-1}(0) \subseteq [\![\mathsf{Bel}_i(\phi)]\!]$. Let $w \in k^{-1}(0)$ (such w must exist by Definition 1). Since $w \in [\![\mathsf{Bel}_i(\phi)]\!]$, $K(i,w)([\![\phi]\!] \geq \mathsf{b}$. But, by the admissibility requirement on commitment states, $k = K(i,w)$. So $k([\![\phi]\!]) \geq \mathsf{b}$. □

In addition, we have the converse of Fact 18: Declarative update with φ ensures that assertoric commitment to $\mathsf{Bel}_i(\varphi)$, which follows directly from the requirement that $\mathsf{a} > \mathsf{b}$:

Fact 20 (Assertoric commitment implies belief commitment)**.** *For all* $C_i, \varphi : C_i + \varphi \vDash \mathsf{Bel}_i(\varphi)$.

With these two results in place, we can show that our system accounts for **Moore's paradox**.

Corollary 21 (Moore's paradox explained)**.** *For any commitment state* C_i*:*

(a) $C_i + (\neg\phi \wedge \mathsf{Bel}_i(\phi)) = \bot$ *(b)* $C_i + (\phi \wedge \mathsf{Bel}_i(\neg\phi)) = \bot$ *(c)* $C_i + (\phi \wedge \neg\mathsf{Bel}_i(\phi)) = \bot$

Proof. (a) By Fact 15, it is sufficient to show that $C_i + \neg\phi + \mathsf{Bel}_i(\phi) = \bot$. Suppose otherwise, i.e. that there is k such that $[C_i + \neg\phi + \mathsf{Bel}_i(\phi)](k) = 1$. Clearly, it must be that $\beta_k([\![\neg\phi]\!]) = \geq \mathsf{a} > 0$ and, by Lemma 19, it must be that $\beta_k([\![\phi]\!]) \geq \mathsf{b} > 0$. But (Fact 3), a belief function can assign positive rank to at most one of $[\![\phi]\!]$ and $[\![\neg\phi]\!]$. Contradiction. (b) Analogous. (c) Again, it is sufficient to show that $C_i + \phi + \neg\mathsf{Bel}_i(\phi) = \bot$. By Fact 20, $C_i + \phi \vDash \mathsf{Bel}_i(\phi)$. Since $[\![\mathsf{Bel}_i(\phi)]\!] \cap [\![\neg\mathsf{Bel}_i(\phi)]\!] = \emptyset$, by Fact 14, $C_i + \phi + \neg\mathsf{Bel}_i(\phi) = \bot$. □

The cases (a-c) in this corollary correspond to the Moorean sentence we have been working with (2) and the two readings of the classical Moore-sentence, (18) and (19):

(2)	It is not raining but I believe it is.	Corollary (21a)
(18)	It is raining but I believe it is not raining.	Corollary (21b)
(19)	It is raining but it is not the case that I believe it is raining.	Corollary (21c)

5 Conclusion

In this paper, I have combined a ranking-theoretic semantics for *believe* with an update-semantics style commitment-based account of declarative force. I have shown that the resulting system compositionally accounts for Moore's paradox, and at the same time can account for the fact that *I believe* functions as a hedge, while predicting a number of other desiderata.

It is noteworthy that the factor that motivated me to employ ranking theory is the desideratum **closure**. I still think that **closure** is desirable, and so it is encouraging to see that the two main phenomena of interest can be jointly accounted for with a theory of ranked belief that delivers on **closure**. However, committed Bayesians may be willing to jettison the principle.

The good news for such a Bayesian is that the commitment-based account as developed here can be combined with a probabilistic threshold account, in a way that predicts all of our desiderate except **closure**. Here is how: Replace the function K in our models with a function $\mathbb{P}$ that assigns to each agent-world pair a probability distribution, and replace b by a threshold $\theta_{\mathsf{b}} > 0.5$. Admissibility then amounts to the following: A probability distribution P is admissible for i iff for all worlds w in the *support* of $P : \mathbb{P}(i, w) = P$, and a model is admissible iff $\mathbb{P}$ only assigns admissible probability distributions. Commitment states become truth functions of admissible probability distributions, the assertion threshold is $\theta_{\mathsf{a}} \in \mathbb{R}$ such that $\theta_{\mathsf{a}} > \theta_{\mathsf{b}}$. The update operation is adjusted in the obvious way: $C_i + \phi = \lambda P.C_i(P)\ \&\ P(\phi) \geq \theta_{\mathsf{a}}$.

What this shows is that the choice of a theory of graded belief can be made on independent grounds. People like me, who find **closure** compelling will find the version of the account proposed in the main text attractive and comforting. Committed Bayesians, being Bayesians, may prefer the account sketched in the previous paragraph. The contribution of the present paper is not an argument for one representation of graded belief over another, but rather in that it shows that Moore's paradox and hedging with *I believe* can be explained in the 'natural ways' sketched at the beginning of the paper, that both can be explained together, and that this can be done compositionally, by combining a graded theory of belief with a commitment-based understanding of declarative force.

References

Condoravdi, C. and Lauer, S.: 2011, Performative verbs and performative acts, *in* I. Reich, E. Horch and D. Pauly (eds), *Sinn and Bedeutung 15*, Universaar, Saarbrücken, pp. 149–164.

Hawthorne, J., Rothschild, D. and Spectre, L.: 2015, Belief is weak, *Philsophical Studies* **173**(5).

Krifka, M.: 2014, Embedding illocutionary acts, *in* T. Roeper and P. Speas (eds), *Recursion, Complexity in Cognition*, Springer, Berlin, pp. 125–155.

Krifka, M.: 2015, Bias in commitment space semantics: Declarative questions, negated quetions, and question tags, *Semantics and Linguistic Theory* **25**, 328–345.

Kyburg, H. E.: 1961, *Probability and the Logic of Rational Belief*, Wesleyan University Press.

Lassiter, D.: 2017, *Graded Modality: Qualitative and Quantitative Perspectives*, Oxford University Press.

Lauer, S.: 2013, *Towards a dynamic pragmatics*, PhD thesis, Stanford University.

Leitgeb, H.: 2014, The review paradox: On the diachronic costs of not closing rational belief under conjunction, *Noûs* **48**(4), 781–793.

Leitgeb, H.: 2015, The Humean thesis on belief, *Aristotelian Society Suppl. Vol.* **89**(1), 143–185.

Lin, H. and Kelly, K. T.: 2012, A geo-logical solution to the lottery paradox, *Synthese* **186**, 531–575.

Makinson, D.: 1965, The paradox of the preface, *Analysis* **25**(6), 205–207.

Raidl, E.: 2017, Completeness for counter-doxa conditionals–using ranked semantics. ms., University of Konstanz.

Raidl, E.: t.a., Ranking semantics for doxastic necessities and conditionals, *The Logica Yearbook 2017*.

Rieger, A.: 2015, Moore's paradox, introspection and doxastic logic, *Thought* **4**, 215–227.

Searle, J. R.: 1969, *Speech Acts: An essay in the philosophy of language*, Cambridge University Press.

Spohn, W.: 1988, Ordinal conditional functions: A dynamic theory of epistemic states, *in* W. Harper and B. Skyrms (eds), *Causation in Decision, Belief Change, and Statistics*, Kluwer, pp. 105–134.

Spohn, W.: 1990, A general non-probabilistic theory of inductive reasoning, *in* R. Shachter, T. Levitt, J. Lemmer and L. Kanal (eds), *Uncertainty in Artificial Intelligence (Volume 4)*, pp. 149–158.

Spohn, W.: 2012, *The laws of belief*, Oxford University Press, Oxford, UK.

Swanson, E.: 2006, *Interactions with context*, PhD thesis, MIT, Cambridge, MA.

van der Hoek, W.: 1993, Systems for knowledge and belief, *Journal of Logic and Computation* **3**(2), 173–195.

Veltman, F.: 1996, Defaults in update semantics, *Journal of Philosophical Logic* **25**, 221–261.

Semantics of metalinguistic focus *

Haoze Li

New York University, New York, USA
haozeli@nyu.edu

Abstract

Focus on metalinguistic aspects of utterances, despite being a robust phenomenon in natural language, is unamenable to the standard semantics of focus. Nonetheless, I show that this type of focus can be understood in terms of standard focus semantics, if we incorporate insights from Pott's [21] multi-dimensional semantics of mixed quotations. Moreover, I develop a scope-taking account to compositionally synthesize focus semantics and quotation semantics.

1 Metalinguistic focus

It is standardly assumed that focusing a word indicates that alternatives to the word are considered. The alternatives are derived from the denotation of the word (Rooth [22, 23]). For example, the focused noun *geese*$_{\text{F}}$ triggers a set of alternatives including all entities of type $e \to t$. However, in natural language, focus is not only determined based on the denotation, i.e., the meaning, of a word, but also based on its *non-semantic aspect*, as exemplified in (1).

(1) a. A: Look! Some gooses are flying.
b. B: No. Some [geese]$_{\text{F}}$ are flying.

Native speakers share the intuition that what B objects to in (1) is not the meaning of A's utterance, but rather the realization of the plural morphology on *goose* A has chosen. Since one's preference of phonological form is an attribute of utterances, focus on these aspects have been said to be 'metalinguistic' (Selkirk [25]).

Metalinguistic focus is a robust phenomenon found not just in English. In the Mandarin example (2), B corrects A's pronunciation of the city name by focusing the final syllable.

(2) a. A: Libai qu-le Ha'erbing.
Libai go-Asp Harbin
'Libai went to Harbing.'

b. B: Ta qu-le Ha'er[bin]$_{\text{F}}$.
he go-Asp Harbin
'He went to Har[bin]$_{\text{F}}$.'

In (2b), focus is assigned to the syllable bin, which does not have an identifiable separate meaning in this case[1]. The focus can only be interpreted in metalinguistic terms.

Metalinguistic focus has the same set of phonological properties as normal, semantic use of focus (Selkirk [25]). The close affinity in ordinary focus and metalinguistic focus naturally leads us to expect a uniform focus theory that accounts for both. However, nontrivial challenges arise when the classical compositional semantics for focus is extended to metalinguistic focus.

The primary difficulty lies in the fact that our classical focus theory is based on denotation, something that phonological forms lack. This property is shared by many semantic theories of focus, like Kratzer [15], Krifka [16] and Rooth [22, 23]. For concreteness I base the discussion

*Thanks to Chris Barker, Simon Charlow, Jess Law, Philippe Schlenker and Anna Szabolcsi.

[1]While most monosyllabic units bear meaning in Mandarin, *Ha'erbin* is a loanword from Manchu and hence it does not have a decomposable meaning for Mandarin native speakers.

on Rooth's focus semantics. In this theory, it is assumed that a focused phrase α_F is associated with two semantic values: the ordinary value $[\![\alpha]\!]$, which is the normal denotation, and the focus value $[\![\alpha]\!]^f$, which is a set of alternative denotations to $[\![\alpha]\!]$. The focus value of a larger phrase Σ embedding α_F is derived from $[\![\alpha]\!]^f$ via the pointwise function application rule. Finally, [23] formalizes his theory of focus licensing by positing an operator $\sim$. This operator requires a contextual restriction C, which is subject to the following constraints:

(3) $$[\![[\Sigma] \sim C]\!] = \begin{cases} [\![\Sigma]\!] \text{ if } C \subseteq [\![\Sigma]\!]^f \wedge [\![\Sigma]\!] \in C \wedge \exists y[y \in C \wedge y \neq [\![\Sigma]\!]] \\ \text{otherwise undefined} \end{cases}$$

It says that C is the subset of $[\![\alpha]\!]^f$ and must contain a contextual salient antecedent that differs from $[\![\alpha]\!]$. Focus in a sentence is felicitous only if the requirement of $\sim$ is satisfied. Although $\sim$ is not present in all versions of compositional semantics for focus, its requirements are always implemented in some way, for example by the **assert** operator in Krifka [16].

Returning to the dialogue in (1), (1b) and its antecedent (1a) have the same denotation, as *gooses* share the same denotation as *geese*, at least from B's perspective. What this means is that (1b) does not semantically differ from its contextual antecedent. As a result, the requirement of $\sim$ is not satisfied and B's use of focus on *geese* should not be appropriate in this context, contrary to fact.

In order to take care of metalinguistic focus, one may try directly upgrading the focus theory with a mechanism that can generate focus alternatives with metalinguistic information. A potential analysis is made possible by Katzir's [13] structural alternative approach. To put simply, metalinguistic focus can be understood as focus on the whole word, which has both form and meaning. The sentence in (1b) can be taken to be a structure consisting of words, i.e., units with form and meaning. Its focus alternative is derived by replacing the focused word *geese* with the word *gooses*. Consequently, although (1b) has the same semantic interpretation as its alternative, i.e., (1a), they have different forms: (1b) has the word *geese*, but its alternative has the word *gooses*.

However, the solution generates an undesirably strong requirement on focus licensing. Since a word has both form and meaning, judging whether focus is licensed should depend not only on meaning but also on form. The condition is too strong as it wrongly rules out the felicitous use of focus in (2b). Following the structural alternative assumption, we can generate alternatives to (2b) by replacing *Ha'erbin* with other nouns. So, the focus alternative set is $\{[_S$ ta shi qu-le $x] \mid x$ is a noun$\}$. The antecedent sentence (2a) cannot belong to this set, because the first word in (2a) is *Libai* instead of *ta*. The requirement of $\sim$ should not be satisfied.

2 The two-dimensional meaning of metalinguistic focus

I analyze metalinguistic focus as **focus on linguistic expressions**. Linguistic expressions have been studied in the literature of quotation, such as Koev [14], Maier [17], Potts [21] and Shan [28]. Based on the previous studies, I enrich the ontology by adding the type of linguistic expressions u. The domain of linguistic expressions D_u contains all possible phonological strings, not only the ones that are a part of the language. This domain is closed under concatenation $^\frown$. If a and b are phonological strings (linguistic expressions), then $\mathsf{a}^\frown\mathsf{b}$ is also a phonological string (linguistic expression). Throughout this paper I write linguistic expressions in sans serif. Other basic types are individuals (type e), truth values (type $t := \{0, 1\}$) and variables ($\mathcal{V} := \{x, y, v, ...\}$). I assume that models $\mathcal{M}$ are tuples of the form $\langle L, D, [\![\cdot]\!]^c\rangle$, where L is a natural language, D is the domain of any type and $[\![\cdot]\!]^c : L \to D$ is the interpretation function

relativized to an utterance context c, which involves the information of the author, the hear, the world and the time of an utterance (Kaplan [11]; Schlenker [24]).

Following Potts [21], I propose a two-dimensional semantics for linguistic expressions. In particular, I define an operator $\ulcorner\cdot\urcorner$ to model the semantic contribution of a linguistic expression. This function is applied to a linguistic expression u and returns a pair involving the meaning of u in the context c and an 'expression' meaning, as in (4).

(4) $[\![\ulcorner \mathsf{u} \urcorner]\!]^c = \begin{cases} (\!|\mathsf{u}|\!)(c) \bullet \textbf{exp}(c, \mathsf{u}, (\!|\mathsf{u}|\!)(c)) & \text{if } \mathsf{u} \text{ is a meaning bear element in } c; \\ \text{otherwise, undefined} \end{cases}$

a. c is an utterance context.

b. $(\!|\cdot|\!)$ is a function taking a linguistic expression u and returning another function from an utterance context c to the content that u is used to express in c (i.e., Kaplanian characters, see also Shan [28])

c. **exp** is a three-place predicate, associating a context and a linguistic expression to a semantic representation (cf. Maier [17]):

$$\textbf{exp}(c, \mathsf{u}, x) ::= \text{the linguistic expression } \mathsf{u} \text{ is used to express } x \text{ in } c$$

d. $\alpha \bullet \beta$ stands for $\langle \alpha, \beta \rangle$

The 'expression' meaning captures the implication of using metalinguistic focus. For example, in (1b), the core proposition is that there are some geese flying. B also indicates that the intended property is expressed by the phonological form geese, instead of gooses (see also Bolinger [4]). The 'expression' meaning conveys a non-at-issue information. It is not canceled when the sentence is embedded under a truth value negation[2] or a modal. For example, in (5a), *it's not true* only negates the at-issue meaning of its complement. Therefore, the continuation, which confirms that the speaker managed to solve the problem, is not felicitous. In (5b), it is clear that the information of the plural form of *mongoose* project out of the scope of *might*.

(5) a. ?*It's not true that I [mi͡ʸənɪ̆jd]$_{\text{F}}$ to solve the problem — I [mæ͡ʸənɪ̆jd]$_{\text{F}}$ to solve the problem. (Horn [10]: 146)

b. Yesterday, Lee might have caught two mongeese. — Uh, sorry, he might have caught two mon[gooses]$_{\text{F}}$.

The multi-dimensional semantics was designed by Karttunen and Peters [12] to compositionally derives the non-at-issue meaning. This approach has recently been revived by Dekker [7], Gutzmann [8], Potts [20] and so on.

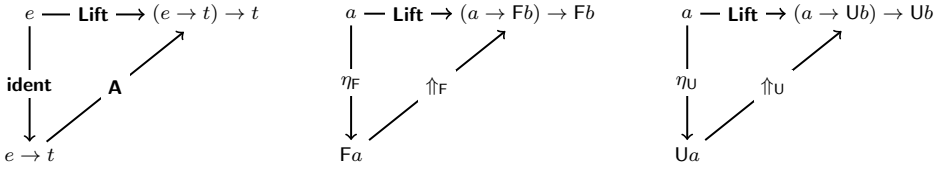
Returning to our examples, for instance (1b), if focus is assigned to the linguistic expression geese, then the puzzle can potentially be resolved. The salient alternative to geese is another linguistic expression gooses in the context. Applying the $\ulcorner\cdot\urcorner$ to them yields:

(6) a. $[\![\ulcorner \text{geese} \urcorner]\!]^c = \lambda x.\textbf{*goose}(x) \bullet \textbf{exp}(c, \text{geese}, \lambda x.\textbf{*goose}(x))$

b. $[\![\ulcorner \text{gooses} \urcorner]\!]^c = \lambda x.\textbf{*goose}(x) \bullet \textbf{exp}(c, \text{gooses}, \lambda x.\textbf{*goose}(x))$

As a result, the two linguistic expressions give rise to different 'expression' meanings, though their semantic values are identical relative to c: the semantic value of geese is the property $\lambda x.\textbf{*goose}(x)$ in the B's (the author of c) English and gooses denotes the same thing in the A's (the hearer of c) English. Then, if we can compositionally derive the meaning of (1b) and make the 'expression' meaning project globally, the sentence *some* $\ulcorner$*geese*$\urcorner$ *are flying* does not have the same denotation as *some* $\ulcorner$*gooses*$\urcorner$ *are flying*. In this sense, the problem discussed in section

[2] It is noticed that the 'expression' meaning can be negated by metalinguistic use of *not* (Horn [10]). In the literature, it is still debated whether metalinguistic negation is one use of *not* or the homonym of *not*.



(a) Partee Triangle (b) **Lift**, η_{F}, $\Uparrow_{\mathsf{F}}$ (c) **Lift**, η_{U}, $\Uparrow_{\mathsf{U}}$

Figure 1: Decomposition of **Lift**

1 does not arise. In the next section, I will show that the goal sketched here can be achieved by fully composing the linguistic expression that $\ulcorner\cdot\urcorner$ operates on with other ordinary elements. In my compositional analysis, only linguistic expressions give rise to the information of form via the application of $\ulcorner\cdot\urcorner$. Consequently, the forms of other ordinary items do not affect the semantic computation of focus licensing. We will not run into the under-generation problem as the structural alternative approach faces.

3 Metalinguistic focus takes scope

In this section, I show that metalinguistic focus can be composed with other ordinary lexical items. The main idea follows the classical Montagovian method of composing quantifiers, i.e., metalinguistic focus, which has a 'fancy' type, **takes scope out of a bigger constituent**, which has a 'plain' type. These two constituents then compose with the help of two pairs of type shifters and Function Application.[3] Note that there is an alternative 'in-situ' compositional analysis, which is more faithful to Rooth's focus semantics and Pott's multi-dimensional semantics. I will return to this analysis in section 5 and show that it has some non-trivial problems.

Partee [18] defines a group of type shifters, **Lift**, **A** and **ident**, that connect basic NP types e, $e \to t$ and $(e \to t) \to t$. Their relationship is summarized in the famous Partee triangle, as shown in Figure 1a. The diagram fully commutes. Hence, **Lift** can be considered as the composition $\mathbf{ident} \circ \mathbf{A}$.

There is no real reason to assume that the idea of type shifting only manifests itself in the nominal domain. **Lift** as a polymorphic function is defined for arbitrary input types (Hendricks [9]; Partee and Rooth [19]). Moreover, following Partee's logic, Charlow [6] shows that **Lift** can be decomposed in other ways. Hence, we may have various groups of type shifters, which are used to compose the items with 'fancy' types, like alternatives or pairs, with ordinary items.

This approach can be extended to focus. Following Rooth [22, 23], I assume that the focused phrase α_{F} denotes a pair consisting of its ordinary value $[\![\alpha]\!]$ and the alternative set to $[\![\alpha]\!]$. If α has some type a, then the type of α_{F} is $a \times (a \to t)$. This 'fancy' type is abbreviated as $\mathsf{F}a$. I define the type shifting functions η_{F} and $\Uparrow_{\mathsf{F}}$ in (7) (see also Charlow [5]; Shan [26]).

(7) a. $\eta_{\mathsf{F}}(x) := x \bullet \{x\}$ $\qquad \mathbf{F} : a \to \mathsf{F}a$

b. $(x \bullet X)^{\Uparrow_{\mathsf{F}}} := \lambda f.\mathbf{fst}(f(x)) \bullet \bigcup_{x' \in X} \mathbf{snd}(f(x'))$ $\qquad \Uparrow_{\mathsf{F}} : \mathsf{F}a \to ((a \to \mathsf{F}b) \to \mathsf{F}b)$

fst and **snd** are operators on pairs. They yield the first member and second member of a pair,

[3] The compositional mechanism used here comprises something known to Category theorists and computer scientists as a 'monad.' I will not formally introduce monads in this paper, but rather represent the monadic spirit in a way that linguists are more familiar with (see also Charlow [6]).

respectively. Through η_{F}, any value can be mapped in a consistent way to a paired value, with the first member the input value and the second member a singleton containing the input value. $\Uparrow_{\mathsf{F}}$ allows an item bearing focus to take scope. Applying the two functions to x, i.e., $(\eta_{\mathsf{F}}(x))^{\Uparrow_{\mathsf{F}}}$, we have actually lifted x from a to $(a \to \mathsf{F}b) \to \mathsf{F}b$. This is essentially **Lift**. Based on this connection, we can draw a Partee-style triangle for a type-a item and its type shifters, as shown in Figure 1b.

Figure 2a shows a sample derivation of *some [geese]*$_{\mathrm{F}}$ *are flying* with the use of the type shifters. The focused phrase takes scope via the application of $\Uparrow_{\mathsf{F}}$, as in (8a). In this paper, I present the scope-taking mechanism using Quantifier Raising. It is also compatible with other theories of scope-taking, such as Continuation (Barker and Shan [2]), Flexible Types (Hendricks [9]), etc. Applying the function η_{F} to *some* P *are flying* results in a pair of the proposition $\exists x.P(x) \wedge \mathbf{fly}(x)$ and a singleton set containing it, as in (8b). The final result is given in (8c).

(8) a. $(\llbracket \text{geese} \rrbracket^c \bullet \mathbf{alt}(\llbracket \text{geese} \rrbracket^c))^{\Uparrow_{\mathsf{F}}} = \lambda f.\mathbf{fst}(f(\llbracket \text{geese} \rrbracket^c)) \bullet \bigcup_{P \in \mathbf{alt}(\llbracket \text{geese} \rrbracket^c)} \mathbf{snd}(f(P))$

b. $\eta_{\mathsf{F}}(\exists x.P(x) \wedge \mathbf{fly}(x)) = \exists x.P(x) \wedge \mathbf{fly}(x) \bullet \{\exists x.P(x) \wedge \mathbf{fly}(x)\}$

c. $(\llbracket \text{geese} \rrbracket^c \bullet \mathbf{alt}(\llbracket \text{geese} \rrbracket^c))^{\Uparrow_{\mathsf{F}}}\ \lambda P.\ \eta_{\mathsf{F}}(\exists x.P(x) \wedge \mathbf{fly}(x))$
$= \exists x.\textbf{*goose}(x) \wedge \mathbf{fly}(x) \bullet \{\exists x.P(x) \wedge \mathbf{fly}(x) \mid P \in \mathbf{alt}(\llbracket \text{geese} \rrbracket^c)\}$

As discussed in section 2, the linguistic expression u operated by $\ulcorner \cdot \urcorner$ denotes a pair consisting of the denotation of u in the context c, i.e., $(\!|\mathsf{u}|\!)(c)$, and the propositional 'expression' meaning, i.e., u is used to express $(\!|\mathsf{u}|\!)(c)$ in c. If $(\!|\mathsf{u}|\!)(c)$ has the type a, then $\ulcorner \mathsf{u} \urcorner$ has the type $a \times t$, which is abbreviated as $\mathsf{U}a$. Along the same lines as focus, we can define another pair of type shifters, as in (9), to integrate $\ulcorner \mathsf{u} \urcorner$ into compositional semantics.

(9) a. $\eta_{\mathsf{U}}(x) := x \bullet \mathbf{T}$ $\qquad \eta_{\mathsf{U}} : a \to \mathsf{U}a$

b. $(x \bullet p)^{\Uparrow_{\mathsf{U}}} := \lambda f.\mathbf{fst}(f(x)) \bullet p \wedge \mathbf{snd}(f(x))$ $\qquad \Uparrow_{\mathsf{U}}\!: \mathsf{U}a \to ((a \to \mathsf{U}b) \to \mathsf{U}b)$

Similar to η_{F}, η_{U} maps any value to a trivial pair value (being paired with the tautology $\mathbf{T}$). $\Uparrow_{\mathsf{U}}$ is a mapping from pairs into pair-friendly scope takers. The composition $\Uparrow_{\mathsf{U}} \cdot \eta_{\mathsf{F}}$ is also **Lift**. Their relationship is demonstrated in Figure 1c, another Partee-style triangle.

Returning to (1b), I analyze it as (10): the linguistic expression geese inside $\ulcorner \cdot \urcorner$ bears focus.

(10) Some $\ulcorner \text{geese}_{\mathrm{F}} \urcorner$ are flying.

The LF depicting the derivation of (10) is given in Figure 2b. The derivation consists of two steps. First, $\ulcorner \text{geese}_{\mathrm{F}} \urcorner$ takes scope via the application of $\Uparrow_{\mathsf{U}}$. η_{U} is applied to its scope, i.e., $\exists x.P(x) \wedge \mathbf{fly}(x)$. Second, $\text{geese}_{\mathrm{F}}$ as a focused item also takes scope, leaving a type u trace inside $\ulcorner \cdot \urcorner$. η_{F} is applied to its scope, in which $\ulcorner \mathsf{u} \urcorner$ composes with $\exists x.P(x) \wedge \mathbf{fly}(x)$ through $\Uparrow_{\mathsf{U}}$ and η_{U}, as shown in (11).

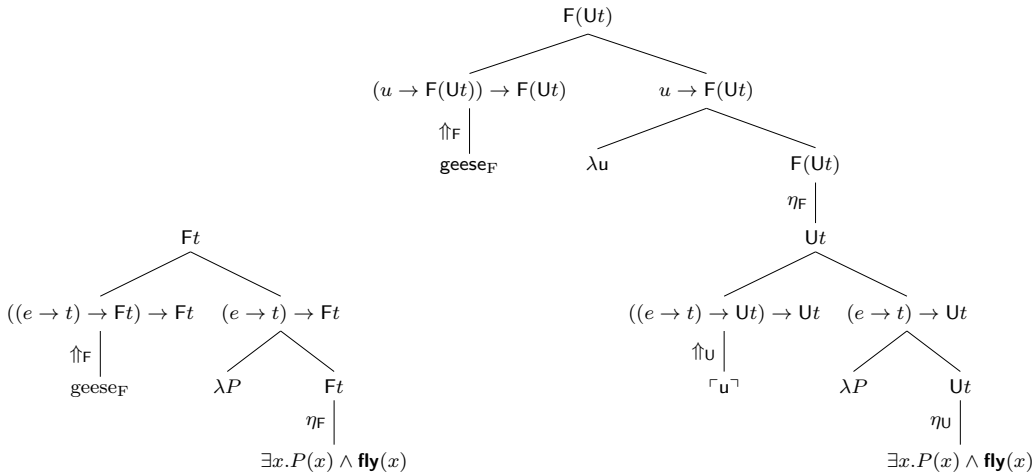
(11) a. $(\llbracket \ulcorner \mathsf{u} \urcorner \rrbracket^c)^{\Uparrow_{\mathsf{U}}} = ((\!|\mathsf{u}|\!)(c) \bullet \mathbf{exp}(c, \mathsf{u}, (\!|\mathsf{u}|\!)(c)))^{\Uparrow_{\mathsf{U}}} = \lambda f \begin{pmatrix} \mathbf{fst}(f((\!|\mathsf{u}|\!)(c))) \\ \bullet \\ \mathbf{exp}(c, \mathsf{u}, (\!|\mathsf{u}|\!)(c)) \wedge \mathbf{snd}(f((\!|\mathsf{u}|\!)(c))) \end{pmatrix}$

b. $\eta_{\mathsf{U}}(\exists x.P(x) \wedge \mathbf{fly}(x)) = \exists x.P(x) \wedge \mathbf{fly}(x) \bullet \mathbf{T}$

c. $((\!|\mathsf{u}|\!)(c) \bullet \mathbf{exp}(c, \mathsf{u}, (\!|\mathsf{u}|\!)(c)))^{\Uparrow_{\mathsf{U}}}\ \lambda P.\ \eta_{\mathsf{U}}(\exists x.P(x) \wedge \mathbf{fly}(x)) = \begin{matrix} \exists x.(\!|\mathsf{u}|\!)(c)(x) \wedge \mathbf{fly}(x) \\ \bullet \\ \mathbf{exp}(c, \mathsf{u}, (\!|\mathsf{u}|\!)(c)) \end{matrix}$

Then, $\text{geese}_{\mathrm{F}}$ is composed with (11c) in the same way as the one illustrated in (12):

(12) a. $(\llbracket \text{geese} \rrbracket^c)^{\Uparrow_{\mathsf{F}}} = \lambda f.\mathbf{fst}(f(\text{geese})) \bullet \bigcup_{\mathsf{u}' \in \mathbf{alt}(\text{geese})} \mathbf{snd}(f(\mathsf{u}'))$



(a) Focus (b) Metalinguistic focus

Figure 2: Composition: Focus and metalinguistic focus take scope

$$\text{b.}\quad \eta_{\mathsf{F}}\begin{pmatrix}\exists x.(\!|\mathsf{u}|\!)(c)(x)\wedge \mathbf{fly}(x)\\ \bullet\\ \mathbf{exp}(c,\mathsf{u},(\!|\mathsf{u}|\!)(c))\end{pmatrix}=\begin{pmatrix}\exists x.(\!|\mathsf{u}|\!)(c)(x)\wedge \mathbf{fly}(x)\\ \bullet\\ \mathbf{exp}(c,\mathsf{u},(\!|\mathsf{u}|\!)(c))\end{pmatrix}\bullet\left\{\begin{array}{c}\exists x.(\!|\mathsf{u}|\!)(c)(x)\wedge \mathbf{fly}(x)\\ \bullet\\ \mathbf{exp}(c,\mathsf{u},(\!|\mathsf{u}|\!)(c))\end{array}\right\}$$

$$\text{c.}\quad (\llbracket\mathsf{geese}\rrbracket^{c})^{\Uparrow_{\mathsf{F}}}\ \lambda\mathsf{u}.\ \eta_{\mathsf{F}}\begin{pmatrix}\exists x.(\!|\mathsf{u}|\!)(c)(x)\wedge \mathbf{fly}(x)\\ \bullet\\ \mathbf{exp}(c,\mathsf{u},(\!|\mathsf{u}|\!)(c))\end{pmatrix}$$
$$=\begin{pmatrix}\exists x.(\!|\mathsf{geese}|\!)(c)(x)\wedge \mathbf{fly}(x)\\ \bullet\\ \mathbf{exp}(c,\mathsf{geese},(\!|\mathsf{geese}|\!)(c))\end{pmatrix}\bullet\left\{\begin{array}{c|l}\exists x.(\!|\mathsf{u}'|\!)(c)(x)\wedge \mathbf{fly}(x) & \\ \bullet & \mathsf{u}'\in\mathbf{alt}(\mathsf{geese})\\ \mathbf{exp}(c,\mathsf{u}',(\!|\mathsf{u}'|\!)(c)) & \end{array}\right\}$$

After these two steps, we apply Rooth's focus interpretation operator $\sim$, which is re-defined in (13) to be compatible with the present analysis. Compositionally, it is an operation for discharging focus values (bringing $\mathsf{F}a$ back to a).

$$(13)\qquad (x\bullet X)\sim C:=\begin{cases}x \text{ if } C\subseteq X\wedge x\in C\wedge\exists y[y\in C\wedge y\neq x]\\ \text{otherwise undefined}\end{cases}\qquad \sim C:\mathsf{F}a\to a$$

Given the context, the free variable C contains two members—*some* ⌜***geese***⌝ *are flying* and *some* ⌜***gooses***⌝ *are flying*. The requirement of $\sim$ is fulfilled.

4 Focus below the word level

The present analysis can be extended to another intriguing focus phenomenon, first discussed by Bolinger [3]—focus below the word level. In particular, focus can be realized on a different syllable in a word than the one stress normally falls on. When this happens, the meaning of a sentence is also affected. Artstein [1] represents Bolinger's observation by revising his 'stalagmite' example. Consider the sentences in (14). The stress of the word *stalagmite* is

normally assigned to the second syllable, as in (14a). In this example, the whole word is considered focused. However, the stress can alternatively be assigned to the final syllable of *stalagmite*, as in (14b). Here, only the syllable mite is focused.

(14) a. John only brought home a [stalágmite]$_{F}$ from the cave.
b. John only brought home a stalag[míte]$_{F}$ from the cave.

These two sentences have different truth conditions. Suppose a scenario that John brought home a stalagmite and a rock from the cave, then (14a) is false, but (14b) can be true. This is because in (14b) the alternative to *stalagmite* is restricted to a word which has a similar form, i.e., *stalactite*. Therefore, (14b) only entails that John didn't bring home a stalactite, but he might have brought home anything else.

Focus below the word level is intriguing because the focus bearing elements, mite in this case, often do not have a meaning, just like focus on phonological forms in metalinguistic focus. In order to resolve this puzzle, Artstein [1] postulates a semantic process of phonological decomposition, which assigns denotations to units that lack an independent meaning. Briefly, the focused part of a word denotes a phonological string and the rest of the word is a function from phonological strings to word meanings. For example, $[\![\text{stalag}]\!]([\![\text{mite}]\!]) = [\![\text{stalagmite}]\!]$. Therefore, the meaning of the word parts is fully compositional. Any compositional semantics of focus can apply to parts of words without modification.

Although phonological decomposition is also used to explain focus on phonological strings, it cannot be extended to metalinguistic focus. Let's return to the Mandarin example (2), repeated in (15), in which the final syllable of *Ha'erbin* is focused and in Mandarin, generally, the first syllable of a word is prosodically prominent. It looks similar to focus below the word level.

(15) a. A: Libai qu-le Ha'erbing.
Libai go-Asp Harbin
'Libai went to Harbing.'

b. B: Ta qu-le Ha'er[bin]$_{F}$.
he go-Asp Harbin
'He went to Har[bin]$_{F}$.'

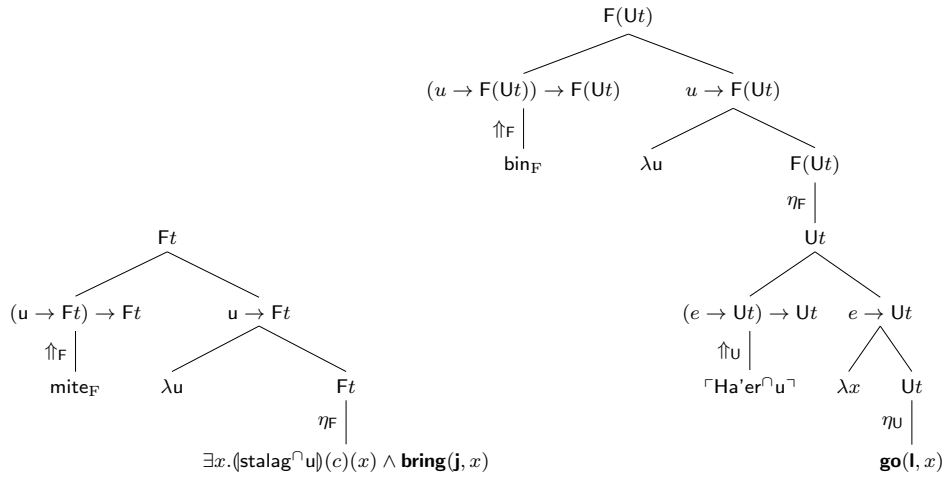
According to phonological decomposition, applying the function denoted by *Ha'er* to the syllable bin yields the denotation of the word, i.e., the capital city of Heilongjiang Province. It cannot resolve the problem pointed out in section 1 and cannot capture the 'expression' meaning of metalinguistic focus.

In fact, focus below the word level is not necessarily metalinguistic (cf. Selkirk [25]). Intuitively, (14b) does not express the contrast on forms. *Stalagmite* contrasts with its alternative *stalactite* with respect to meaning. Additionally, the sentence does not imply the 'expression' meaning. The focus on mite only imposes a restriction on possible alternatives, i.e., their phonological forms must share stalag.

However, the present analysis is able to capture (14) as well as (15). I uniformly assume that focus in both examples is assigned to the part of the linguistic expressions (phonological strings) stalagmite and Ha'erbin, but I apply $\ulcorner\cdot\urcorner$ to Ha'erbin, while $(\!|\cdot|\!)$ to stalagmite. Consequently, $\ulcorner$Ha'erbin$\urcorner$ denotes a pair meaning, whereas $(\!|$stalagmite$|\!)$ denotes a property in the context of utterance. The derivations of (14b) and (15b) are shown in Figure 3a and 3b, respectively.[4] I sketch the composition of (14b) as follows but leave that of (15b) for the reader.

In (14b), the phonological string stalagmite is decomposed into three syllables—sta, lag and mite, which are also linguistic expressions and may compose by concatenation (see section 2). Since mite is focused, it takes scope via the application of $\Uparrow_{F}$. $(\!|\cdot|\!)$ is applied to the remaining parts and transforms a linguistic expression into a character. In the context of utterance c,

[4] I have omited *only* in the derivation for simplification. The classical definition of *only* is compatible with my analysis.



(a) Stalag[mite]$_{\mathrm{F}}$ (b) Ha'er[bin]$_{\mathrm{F}}$

Figure 3: Focus below the word level: Metalinguistic and non-metalinguistic

$(\!|\mathsf{stalag}^{\frown}\mathsf{u}|\!)(c)$ denotes a $e \to t$ type property. So, it directly composes with other lexical items in the sentence, yielding a proposition. Applying η_{F} to the proposition results in (16a). Then, composing the result with $([\![\mathsf{mite}_{\mathrm{F}}]\!])^{\Uparrow_{\mathsf{F}}}$ leads to (16b).

(16) a. $\eta_{\mathsf{F}}(\exists x.(\!|\mathsf{stalag}^{\frown}\mathsf{u}|\!)(c)(x) \wedge \mathbf{bring}(\mathbf{j}, x)) = \begin{matrix} \exists x.(\!|\mathsf{stalag}^{\frown}\mathsf{u}|\!)(c)(x) \wedge \mathbf{bring}(\mathbf{j}, x) \\ \bullet \\ \{\exists x.(\!|\mathsf{stalag}^{\frown}\mathsf{u}|\!)(c)(x) \wedge \mathbf{bring}(\mathbf{j}, x)\} \end{matrix}$

b. $(\mathsf{mite} \bullet \mathbf{alt}(\mathsf{mite}))^{\Uparrow_{\mathsf{F}}}\ \lambda \mathsf{u}.\ \begin{pmatrix} \exists x.(\!|\mathsf{stalag}^{\frown}\mathsf{u}|\!)(c)(x) \wedge \mathbf{bring}(\mathbf{j}, x) \\ \bullet \\ \{\exists x.(\!|\mathsf{stalag}^{\frown}\mathsf{u}|\!)(c)(x) \wedge \mathbf{bring}(\mathbf{j}, x)\} \end{pmatrix}$

$= \begin{matrix} \exists x.(\!|\mathsf{stalag}^{\frown}\mathsf{mite}|\!)(c)(x) \wedge \mathbf{bring}(\mathbf{j}, x) \\ \bullet \\ \{\exists x.(\!|\mathsf{stalag}^{\frown}\mathsf{u}|\!)(c)(x) \wedge \mathbf{bring}(\mathbf{j}, x) \mid \mathsf{u} \in \mathbf{alt}(\mathsf{mite})\} \end{matrix}$

5 Comparison with the 'in-situ' composition

In my analysis, both focus and linguistic expressions take scope. However, scope taking may not be necessary if we only consider simple cases. Let's still take (1b) as an example. Assuming Rooth's focus semantics and Koev's [14] two-dimensional function application, as in (17) (cf. [21]), we can compose the meaning of (1b) without letting the metalinguistic focus take scope.

(17) Two-dimensional function application:
If $[\![\alpha]\!]_{(\sigma\to\tau)\times t} = a_1 \bullet p_2$ and $[\![\beta]\!]_{\sigma\times t} = b_1 \bullet p_2$, then $[\![\alpha(\beta)]\!]_{\tau\times t} = a_1(b_1) \bullet p_1 \wedge p_2$.

When the $\ulcorner\cdot\urcorner$ operator is applied to $\mathsf{geese}_{\mathsf{F}}$, a two-dimensional meaning is generated which interacts focus. As the ordinary value, a two-dimensional meaning is computed based on the form geese. As the focus value, a set of two-dimensional meanings are computed based on the

alternative forms .

(18) a. $[\![\ulcorner \textsf{geese}_{\textsf{F}} \urcorner]\!]^c = \ulcorner \textsf{geese} \urcorner = \lambda x.\textbf{*goose}(x) \bullet \textbf{exp}(c, \textsf{geese}, \lambda x.\textbf{*goose}(x))$
b. $[\![\ulcorner \textsf{geese}_{\textsf{F}} \urcorner]\!]^c_f = \{\ulcorner \textsf{u} \urcorner \mid \textsf{u} \in \textbf{alt}(\textsf{geese})\}$

We also assume that other lexical items have trivial two-dimensional meanings, i.e., they denote pairs of their denotations and $\mathbf{T}$. For example, $[\![\text{fly}]\!]^c = \lambda x.\textbf{fly}(x) \bullet \mathbf{T}$ and $[\![\text{some}]\!]^c = \lambda P \lambda Q \exists x.P(x) \wedge Q(x) \bullet \mathbf{T}$. Hence, their focus values are singleton sets of a pair: $[\![\text{fly}]\!]^c_f = \{[\![\text{fly}]\!]^c\}$ and $[\![\text{some}]\!]^c_f = \{[\![\text{some}]\!]^c\}$. In the ordinary dimension, we compose (1b) via the two-dimensional function application, yielding (19a), the ordinary value of (1b); whereas in the focus dimension, we pointwisely compose $[\![\ulcorner \textsf{geese}_{\textsf{F}} \urcorner]\!]^c_f$ with $[\![\text{some}]\!]^c_f$ and $[\![\text{fly}]\!]^c_f$, yielding (19b), the focus value of (1b). This result is equivalent to (12c).

(19) a. $\exists x.\textbf{*goose}(x) \wedge \textbf{fly}(x) \bullet \textbf{exp}(c, \textsf{geese}, \lambda x.\textbf{*goose}(x))$
b. $\{\exists x.(\!|\textsf{u}|\!)(c)(x) \wedge \textbf{fly}(x) \bullet \textbf{exp}(c, \textsf{u}, \lambda x.(\!|\textsf{u}|\!)(c)(x)) \mid \textsf{u} \in \textbf{alt}(\textsf{geese})\}$

Indeed, the 'in-situ' composition presented here is more in line with the classical version of Rooth's focus semantic, but it also inherits its problems. First, the standard λ-abstraction rule cannot be applied. Consider (20a), whose LF is given in (20b). The quantifier *nobody* takes scope. At the ordinary dimension, if we apply the standard λ-abstraction rule to the scope of *nobody*, then we end up with a type $e \to (t \times t)$ element, which cannot compose with *nobody* by (17) due to type mismatch.

(20) a. The pólice, uh sorry, the [políce]$_{\text{F}}$ arrested nobody.
b. $\underbrace{\text{nobody}}_{o:\ ((e\to t)\to t)\times t \quad f:\ (((e\to t)\to t)\times t)\to t} \quad \lambda x\ \underbrace{[\text{the [políce]}_{\text{F}}\ \text{arrested}\ x]}_{o:\ t\times t \quad f:\ (t\times t)\to t}$

In the focus dimension, λ-abstraction has to apply to the focus alternative set generated by *[políce]*$_{\text{F}}$. Shan [27] has already shown that λ-abstraction over an alternative set is problematic.

Second, the 'in situ' composition predicts that metalinguistic focus and ordinary focus must be evaluated unselectively when they co-occur in the scope of a focus sensitive adverbial like *only*. This is not true. Consider (21), where *only* is associated with *Peter* rather than *coffee*.

(21) Lee only asked [Peter]$_{\text{F}}$ to buy coffin. Uh sorry, he only asked [Peter]$_{\text{F}}$ to buy co[ffee]$_{\text{F}}$.

These two problems do not arise with the scope-taking approach. It has already been shown in other studies (Charlow [5, 6]) that enriched composition coupled with scope-taking is compatible with standard λ-abstraction. In addition, since focus takes scope, the selectivity effect in (21) can be captured (cf. Krifka [16]).

References

[1] Ron Artstein. Focus below the word level. *Natural Language Semantics*, 12(1):1–22, 2004.

[2] Chris Barker and Chung-chieh Shan. *Continuations and natural languages*. Oxford University Press, Oxford, 2015.

[3] Dwight Bolinger. Contrastive accent and contrastive stress. *Language*, 37:83–96, 1961.

[4] Dwight Bolinger. *Intonation and its parts*. Standford University Press, Standford, CA., 1986.

[5] Simon Charlow. *On the semantics of exceptional scope*. PhD thesis, New York University, 2014.

[6] Simon Charlow. The scope of alternatives: Indefiniteness and islands. *Linguistics and Philosophy*, to appear.

[7] Paul Dekker. A multi-dimensional treatment of quantification in extraordinary English. *Linguistics and Philosophy*, 21:101–127, 2008.

[8] Daniel Gutzmann. Compositional multidimensionality and the lexicon-semantics interface. In *Proceedings of the Tenth International Workshop of Logic and Engineering of Natural Language Semantics*, pages 206–219, 2013.

[9] Herman Hendricks. *Studied Flexibility: Categories and Types in Syntax and Semantics*. PhD thesis, University of Amsterdam, Amsterdam, 1993.

[10] Lawrence Horn. Metalinguistic negation and pragmatic ambiguity. *Language*, 61:121–174, 1985.

[11] David Kaplan. Demonstratives: An essay on the semantics, logic, metaphysics, and epistemology of demonstratives and other indexicals. In Joseph Almog, John Perry, and Howard Wettstein, editors, *Themes From Kaplan*, pages 481–564. Oxford University Press, Oxford, 1989.

[12] Lauri Karttunen and Stanley Peters. Conventional implicature. In O.-K. Oh and D. A. Dinneen, editors, *Syntax and Semantics 11: Presupposition*, pages 1–56, New York, 1979. Academic Press.

[13] Roni Katzir. Structurally-defined alternatives. *Linguistics and Philosophy*, 30:669–690, 2008.

[14] Todor Koev. Quotational indefinites. *Natural Language and Linguistic Theory*, 35:367–396, 2017.

[15] Angelika Kratzer. The representation of focus. In Arnim von Stechow and Dieter Wunderlich, editors, *Semantics: An international handbook of contemporary research*, pages 825–834. Walter de Gruyter, Berlin, 1991.

[16] Manfred Krifka. A compositional semantics for multiple focus constructions. In Steven Moore and Adam Zachary Wyner, editors, *Proceedings of the First Semantics and Linguistic Theory Conference*, pages 127–158, Cornell University, 1991. Cornell University Working Papers in Linguistics.

[17] Emar Maier. Mixed quotation: The grammar of apparently transparent opacity. *Semantics & Pragmatics*, 7:1–67, 2014.

[18] Barbara H. Partee. Noun phrase interpretation and type-shifting principles. In D. de Jongh Groenendijk and Martin Stokhof, editors, *Studies in discourse representation theory and the theory of generalized quantifiers*, pages 115–143, 1986.

[19] Barbara Hall Partee and Mats Rooth. Generalized conjunction and type ambiguity. In R. Bäuerle, C. Schwarze, and Arnim von Stechow, editors, *Meaning, Use and Interpretation of Language*, pages 362–383. de Gruyter, 1983.

[20] Christopher Potts. *The Logic of Conventional Implicatures*. Oxford University Press, Oxford, 2005.

[21] Christopher Potts. The dimensions of quotation. In Chris Barker and Pauline Jacobson, editors, *Direct Composition*, pages 405–431, Oxford, 2007. Oxford University Press.

[22] Mats Rooth. *Association with Focus*. PhD thesis, University of Massachusetts, Amherst, 1985.

[23] Mats Rooth. A theory of focus interpretation. *Natural Language Semantics*, 1(1):117–121, 1992.

[24] Philippe Schlenker. A plea for monsters. *Linguistics and Philosophy*, pages 29–120, 2003.

[25] Elisabeth O. Selkirk. *Phonology and syntax : the relation between sound and structure*. MIT Press, Cambridge, Mass., 1984.

[26] Chung-chieh Shan. Monads for natural language semantics. In Kristina Striegnitz, editor, *Proceedings of the ESSLLI 2001 Student Session*, pages 285–298. 13th European Summer School in Logic, Language and Information., 2001.

[27] Chung-chieh Shan. Binding alongside hamblin alternatives calls for variable-free semantics. In Robert B. Young, editor, *Proceedings of SALT XIV*, pages 289–304, Cornell University, Ithaca, NY, 2004. CLC Publications.

[28] Chung-chieh Shan. The character of quotation. *Linguistics and Philosophy*, 33:417–443, 2010.

Implicative inferences and causality in *enough* and *too* constructions*

Prerna Nadathur

Department of Linguistics, Stanford University
pnadathur@stanford.edu

Abstract

This paper proposes a new account of the aspect-dependent implicative behavior of *enough* and *too* constructions (E&T). Against Hacquard (2005)'s claim that E&T are inherently complement-entailing, I propose that they simply attribute a capacity to their subjects, but do not force complement entailment. Actualization under perfective is driven by 'actualistic' aspectual coercion (Homer, 2011), which applies only to a specific set of stative predicates. This aligns perfective E&T with recent treatments of implicatives (Baglini & Francez, 2016; Nadathur, 2016), and opens up a new approach to the longstanding puzzle of actuality entailments on ability modals (Bhatt, 1999).

1 Implicative inferences

Implicative verbs entail the truth of their complements, reversing this entailment under negation. The relationships in (1a-1b) suggest a false equivalence between the matrix and embedded propositions; Karttunen (1971) proposes that this equivalence is precluded by presuppositions of necessity and sufficiency associated with use of the implicative.

(1) a. Morgan managed to solve the riddle. $\vdash$ *Morgan solved the riddle.*
b. Morgan did not manage to solve the riddle. $\vdash$ *Morgan did not solve the riddle.*

Karttunen also identifies a set of predicates that are 'optionally' implicative, in that they defeasibly implicate truth values for their complements. *Enough* and *too* (E&T) constructions are of this type. Like the entailments in (1), the inferences in (2) reverse with matrix negation.

(2) a. Juno was fast enough to win the race. $\rightsquigarrow$ *Juno won the race.*
b. Juno was too slow to win the race. $\rightsquigarrow$ *Juno did not win the race.*

Defeasibility suggests a pragmatic analysis, but this is complicated by the fact that E&T inferences in French are governed by grammatical aspect (Hacquard, 2005). In the imperfective, *être assez rapid/e* patterns with (2a). In the perfective, however, we get full entailment.[1] The same contrast arises with *be too slow* (*être trop lent/e*), *modulo* negation (see 2b).

(3) a. *Juno était assez rapide pour gagner la course, mais elle n'a jamais gagné.*
'Juno was-IMPF fast enough to win the race, but she never won.'
b. *Juno a été assez rapide pour gagner la course, #mais elle n'a pas gagné.*
'Juno was-PFV fast enough to win the race, #but she did not win.'

*Thanks are due to Cleo Condoravdi, Itamar Francez, and Sabine Iatridou for discussion of these ideas. Additional thanks go to Jérémie Wenger for judgements and insight about French aspect.

[1] French marks perfective aspect with the *passé composé*, which is a compound of the auxiliary *avoir*(=*have*) with the past participle of the main verb. For ease of readability, I gloss *passé composé* simply as PFV.

Hacquard proposes that E&T implicativity is not optional: like *manage*, E&T constructions presuppose necessary and sufficient conditions and 'at base' entail their complements. To address defeasibility in (3a), Hacquard appeals to the generic interpretation associated with imperfective aspect (Bhatt, 1999). As she points out, however, imperfective marking does not have the same effect on true implicatives: *réussir*(=manage) entails regardless of aspect.

(4) a. *Juno réussissait à gagner la course, #mais elle n'a jamais gagné.*
'Juno manage-IMPF to win the race, #but she never won.'

b. *Juno a réussi à gagner la course, #mais elle n'a pas gagné.*
'Juno manage-PFV to win the race, #but she didn't win.'

'Optional' implicativity in E&T constructions thus presents two challenges. First, how does aspect derive the entailment contrast in (3a-3b)? Second, if E&T constructions are true implicatives, why are entailments absent (2)-(3a), while *manage* entails across the board? If E&T contructions are not implicatives, what accounts for the perfective entailments?

This paper presents a new proposal for E&T constructions, which addresses these challenges. I take as a starting point the idea that 'true' implicativity is built on relations of causal dependence (Nadathur, 2016). On this analysis, an implicative verb presupposes the existence of a *causally necessary and sufficient* event for the truth of its complement. It asserts whether or not the precondition was satisfied.

I propose that there are two main differences between implicatives and E&T constructions. First, E&T constructions in general presuppose only necessary preconditions. They take on the additional causal sufficiency presupposition of implicatives in a particular subset of cases, when the matrix clause describes an exercisable capacity (e.g., *be fast*). Secondly, while implicative verbs assert the occurrence of a causing event, E&T constructions 'at base' assert only the possibility of this causing event. Given these differences, the inferential patterns in (1)-(4) are predicted straightforwardly by the selectional restrictions of perfective and imperfective aspect.

2 Modality, necessity, and sufficiency

2.1 Degree comparatives with a necessity condition

E&T constructions are analyzed as a type of degree comparative, in which an actual degree is compared to a degree associated with the possibility of the complement proposition (Bierwisch, 1987; Meier, 2003; von Stechow et al., 2004; Schwarzschild, 2008).

(5) a. Juno is fast enough to win the race.
Juno is as fast in the real world as she is in some world where she wins the race.
(Her speed makes winning the race possible.)

b. Juno is too slow to win the race.
Juno is slower in the real world than she is in any world where she wins the race.
(Her speed makes winning the race impossible.)

In the spirit of von Stechow et al.'s analysis, I treat *enough* as an equative combined with a universal modal, and *too* as a comparative operator with an existential modal.

(6) a. $[\![\text{enough}]\!]^w :=$
$\lambda Q_{est}\lambda P_{dest}\lambda x_e.\{d : \forall w' \in \textsc{Acc}(w)[Q(x)(w') \rightarrow P(d)(x)(w')]\} \subseteq \{d : P(d)(x)(w)\}$

b. $[\![\text{too}]\!]^w :=$
$\lambda Q_{est}\lambda P_{dest}\lambda x_e.\{d : \exists w' \in \textsc{Acc}(w)[Q(x)(w') \;\&\; P(d)(x)(w')]\} \subset \{d : P(d)(x)(w)\}$

Positive gradable adjectives relate individuals to degrees on a particular scale. Negative adjectives are defined in opposition to their positive poles.

(7) a. $[\![\text{fast}]\!]^w := \lambda d \lambda x.\textsc{speed}(x)(w) \geq d$
b. $[\![\text{slow}]\!]^w := \lambda d \lambda x.\textsc{speed}(x)(w) < d$

Ignoring tense for the moment, this gives us the following interpretations for (2a)-(2b):[2]

(8) a. Juno be fast enough to win the race $= [\![\text{enough}]\!]^w([\![\text{win-race}]\!])([\![\text{fast}]\!])([\![\text{Juno}]\!])$
$\equiv \{d : \forall w' \in \textsc{Acc}(w)[\text{win}(j)(w') \rightarrow \textsc{speed}(j)(w') \geq d]\} \subseteq \{d : \textsc{speed}(j)(w) \geq d\}$
b. Juno be too slow to win the race $= [\![\text{too}]\!]^w([\![\text{win-race}]\!])([\![\text{slow}]\!])([\![\text{Juno}]\!])$
$\equiv \{d : \exists w' \in \textsc{Acc}(w)[\text{win}(j)(w') \;\&\; \textsc{speed}(j)(w') < d]\} \subset \{d : \textsc{speed}(j)(w) < d\}$

The sets under comparison are all intervals – of the form $[0, n]$ in (8a), and $[n, \infty)$ in (8b). We compare maximal elements in (8a): the maximum speed that Juno has in the real world is at least as great as the maximal speed she has in the race-winning world where she is slowest. Her speed thus makes winning the race possible. (8b) compares miminal elements: the minimum degree of speed that Juno does not have is less than the minimum degree she has in any of the race-winning worlds. Thus, Juno needs to be faster than she is for winning to be possible.

We would like to rule out contexts where it is *a priori* impossible that Juno wins, since this renders (2a-2b) infelicitous. Under the present semantics, this situation will return true from (8a) and (8b). We take *enough* and *too* to presuppose that there is at least one accessible world in which Juno wins the race. Since she has a speed in every world, this is equivalent to (9).

(9) $\exists d : \forall w' \in \textsc{Acc}(w)[Q(x)(w') \rightarrow P(d)(x)(w')]$

Since it is nonempty, $\{d : \forall w' \in \textsc{Acc}(w)[Q(x)(w') \rightarrow P(d)(x)(w')]\}$ has a maximum element, which is the minimum speed that Juno must have if it is possible for her to win. Condition (9) therefore mandates the existence of a degree of speed that is necessary for the realization of the E&T complement; it does not, however, mandate the existence of a sufficient condition.

2.2 Modal flavour

Meier (2003) suggests that implicative inferences for English E&T constructions can be explained by the choice of accessibility relation.[3] For instance, (10) asserts Amira's age is asserted to be such that it is possible for her to drink legally. But since we do not assume that people necessarily act on their legal abilities, no inference is predicted.

(10) Amira was old enough to drink. $\not\rightsquigarrow$ *Amira drank.*

(11) a. Morgan was clever enough to solve the riddle. $\rightsquigarrow$ *Morgan solved the riddle.*
b. Morgan was not clever enough to solve the riddle. $\rightsquigarrow$ *Morgan solved the riddle.*

On the other hand, (11) selects a circumstantial accessibility relation, which guarantees that worlds are self-accessible. Meier's idea is that we get implicative inferences just in case the context selects for a *totally realistic* accessibility relation, which only contains the actual world. In order for solving the riddle to be possible in this case, it must take place in the real world, deriving the implicative reading of (11a). This won't work in general, because the same presupposition holds for (11a) and its negation. For (11b) to be felicitous, Morgan must solve the riddle in the real world. But this contradicts the inference we wish to derive.

[2] These semantics derive the duality *enough* and *too*. I therefore focus on *enough* in what remains.
[3] For the purposes of this paper, I make the simplifying assumption that different modals are simply represented by different accessibility relations, rather than spelling out the full apparatus of Kratzer (1981).

2.3 Necessary and sufficient conditions

As things stand, we cannot account for the perfective data. (3b), like (8a), establishes that Juno's speed made it possible for her to win, but does not necessitate that she actually won.

(3b) *Juno a été assez rapide pour gagner la course, #mais elle n'a pas gagné.*
'Juno was-PFV fast enough to win the race, #but she did not win.'

We will derive the complement entailment if the necessary condition on Juno's win is also a sufficient one – that is, if meeting the necessary condition is enough to guarantee winning. Hacquard (2005) therefore replaces (9) with (12), which presupposes that there is a unique degree d that is both necessary and sufficient for the realization of the E&T complement.

(12) $\iota d : \forall w' \in \textsc{Acc}(w)[Q(x)(w') \leftrightarrow P(d)(x)(w')]$

The necessary condition, as above, is represented by the minimum degree d_{nec} such that $\exists w' \in \textsc{Acc}(w) : \textsc{adj}(x)(w') \geq d \;\&\; Q(x)(w')$. The sufficient condition is given by the minimum degree d_{suff} such that $\forall w' \in \textsc{Acc}(w) : \textsc{adj}(x)(w') \geq d_{\text{suff}} \rightarrow Q(x)(w')$. In general, we have $d_{\text{nec}} \leq d_{\text{suff}}$, so (12) guarantees $d_{\text{nec}} = d_{\text{suff}}$. Thus, (6a) reduces to (13), and (8a) to (14).

(13) $[\![\text{enough}]\!]^w := \lambda Q \lambda P \lambda x.P(\iota d : \forall w' \in \textsc{Acc}(w)[Q(x)(w') \leftrightarrow P(d)(x)(w')])(x)(w)$

(14) Juno be fast enough to win the race.
$\equiv \textsc{speed}(j)(w) \geq (\iota d : \forall w' \in \textsc{Acc}(w)[\text{win}(j)(w') \leftrightarrow \textsc{speed}(j)(w') \geq d])$

As long as Acc is reflexive, the entailment follows. None of this is changes for Hacquard with the addition of perfective aspect, which she treats simply as existential closure over time.

The interesting case is the imperfective (3a). Following Bhatt (1999), Hacquard associates imperfective aspect with a genericity operator. GEN quantifies over normal worlds, and the presuppositions get pushed into GEN's restriction (Schubert & Pelletier, 1989). From this we get that Juno has the necessary and sufficient speed to win the race in all normal worlds where such a speed exists, but we cannot draw any conclusions about the real world.

(15) $[\![\textsc{gen}]\!]^w := \lambda Q_{st}[\forall w' \in \textsc{Norm}(w)Q(w')]$

(16) GEN(Juno be fast enough to win the race)
$\equiv \forall w \in \textsc{Norm}(w^*)[(\iota d : \text{win}(j)(w) \leftrightarrow \textsc{speed}(j)(w)) \rightarrow \textsc{speed}(j)(w) \geq d] \,]$

Hacquard (2005) points out that this analysis aligns E&T constructions with implicative verbs: like implicatives, E&T constructions presuppose the existence of a necessary and sufficient condition for their complements, and inform us as to whether or not this condition was met.[4]

The main difference between implicatives and E&T constructions is the modal component. E&T constructions, as we have seen, can appeal to deontic and circumstantial modalities (as well as to epistemic ones), but implicative verbs seem to be restricted to circumstantial modality. This difference, however, does not predict a difference with respect to entailment under imperfective marking: if GEN/IMPF suspends complement entailments for circumstantial E&T, it should do the same for implicatives. But, as (4a) shows, this is not the case.

(17) GEN(Juno manage to win the race)
$\forall w' \in \textsc{Norm}(w)[\exists Q_{set} : [Q(j)(w') \leftrightarrow \text{win-race}(j)(w')] \rightarrow Q(j)(w')]$

(4a) *Juno réussissait à gagner la course, #mais elle n'a jamais gagné.*
'Juno managed-IMPF to win the race, #but she never won.'

[4] This aligns with Karttunen (1971)'s original proposal, though it deviates from the later, 'standard' account (see Karttunen & Peters, 1979). *Modulo* a causal component, the proposal in Nadathur (2016) shares the structure of Karttunen (1971) and Hacquard (2005).

2.4 The sufficiency problem

As given in (12), the E&T sufficiency condition faces two problems: in deontic cases, it is too strong, but in circumstantial cases it is (conceptually) not strong enough!

Consider how presupposition (12) is realized in (10):

(10) Amira was old enough to drink.
presupposes: $\iota d : \forall w \in \textsc{Deon}(w^*)[\text{drive}(J)(w) \leftrightarrow \textsc{age}_w(J) \geq d]$

(10) presupposes that there is an age d such that, in all worlds where the laws are like ours, and are not violated, being d-old necessitates that Amira drinks. This can be satisfied in two ways. Either the context establishes that the only thing holding her back is the legal issue, or there is a law that requires one to drink after reaching a certain age. In the first case, Amira will drink in the real world. The second case never occurs. Consequently, (10) is predicted to be felicitous just in case Amira drinks in the real world. This prediction is incorrect:

(18) *Amira a été assez grande pour boire de l'alcool, mais elle n'a l'a jamais bu.*
'Amira was-PFV old enough to drink alcohol, but she never drank it.'

The problem extends to any E&T construction with a deontic flavour.

Next consider (19). The necessity presupposition was not enough to guarantee entailment. Adding (12) fixed this, but – intuitively speaking – it should not have done so!

(19) *Juno a été assez rapide pour gagner la course.*
presupposes: $\iota d : \forall w \in \textsc{Circ}(w^*)[\text{win}(J)(w) \leftrightarrow \textsc{speed}_w(J) \geq d]$

The problem is this: *being d-fast*, unlike *being d-old*, can be *latent*. *Being d-fast* involves having the capacity to do things at speeds of at least d, but does not require a manifestation of the speed. Clearly, however, there is no speed d such that simply having the capacity to do things d-fast will guarantee a win. Being d-fast can only ensure Juno's success in the event that she exercises her speed. The problem of (19) generalizes to any other exercisable capacity: *be brave enough*, *be strong enough*, *be loud enough*, etc.

In summary: E&T constructions cannot uniformly carry sufficiency presuppositions, since this makes the wrong predictions for the deontic case. On the other hand, a sufficiency presupposition is needed to derive the entailments in (3b), but this must specifically presuppose that *manifesting* the necessary speed d is a sufficient condition for P.

3 Proposal

The discussion in §2 motivates the following proposal:

(20) Let S be a proposition of the form $S = x$ *be* ADJ *enough to* Q, where x is an individual, ADJ a relation between individuals and degrees, and Q a property of individuals. Evaluated with respect to a world w:

(I) S presupposes a degree d_{nec} that is necessary for the possibility of $Q(x)$:

$$\exists d_{\text{nec}} : \forall w' \in \textsc{Acc}(w)[\textsc{adj}(x)(w') < d_{\text{nec}} \rightarrow \neg Q(x)(w')]$$

(II) S asserts that x has least d_{nec} of ADJ in w:

$$[\![S]\!]^w = \textsc{adj}(x)(w) \geq d_{\text{nec}}$$

(III) When ADJ represents an exercisable capacity, S backgrounds:

$$\forall w' \in \textsc{Acc}(w)[\textsc{do-adj}(x)(d_{\text{nec}})(w') \rhd_{\textsc{caus}} Q(x)(w')]$$

where $\textsc{do-adj}(x)(d)(w)$ is a manifestation of d-ADJ by x in w, and $\rhd_{\textsc{caus}}$ is the causal sufficiency operator.[5]

Claims (20.I-II) are just the semantics established in §2.1. The presupposition (20.I) is equivalent to (9).[6] As per §2.4, we omit Hacquard (2005)'s sufficiency presupposition. Claim (20.III) establishes the causal relation that is backgrounded by ADJ, when ADJ represents an exercisable capacity.[7] It follows from (20.I-III) that S will behave like a true implicative utterance just in case ADJ is an exercisable capacity, and S entails DO-ADJ(x)(d) for some degree d.

Proposal (20) captures the inferential behavior in (2)-(3), as well as the non-implicativity of examples like (10) and (18). The French aspectual contrast arises only with exercisable capacities, and is explained in terms of the contrast between latent (stative) attributions of d-adj and an eventive manifestation $\textsc{do-adj}(x)(d)$. The difference between implicative verbs and E&T constructions is attributed to the activation of the sufficiency presupposition (20.III): implicatives are always eventive, while E&T constructions are not.

3.1 Sufficient cause

Proposal (20) establishes the connection between E&T constructions and implicative verbs. On this analysis, E&T constructions do exhibit *true* implicativity, insofar as complement entailments (where they arise) are derived by the same underlying causal structure.[8] Developing an analysis that aligns with that of implicatives is a strong motive for postulating a causal component in the semantics of E&T constructions, but it does not inherently provide evidence for the claim. In this section, I present three pieces of evidence that support to the causal analysis.

Examples like (2) are easily reconciled with a causal interpretation: being fast is naturally understood as a cause of winning races. That this is the correct interpretation of the matrix-complement connection is supported by the oddness of an example like (21).

(21) (?)Juno was loud enough to win the race.

Here, the E&T complement describes a capacity that is not usually involved in race-winning. Making sense of (21) forces us to imagine a chain of causation leading from being loud to race-winning. If, for instance, we imagine that the only runner faster than Juno is very sensitive to noise, and so Juno might force her to slow down by screaming, (21) is improved.

The same point is reinforced when we modify E&T constructions with *because*-clauses:

(22) a. ?Because it was cheap, Morgan was smart enough to buy the ring.
b. Because it was empty, Marie was strong enough to lift the fridge.

[5] I refer to Baglini & Francez (2016), Nadathur (2016) for formal definitions of causal necessity and sufficiency; the *causal dynamics* framework (Schulz, 2011) used for implicatives can be imported for E&T as well.

[6] $\exists d : \forall w' \in \textsc{Acc}(w)[Q(x)(w') \rightarrow \textsc{adj}(x)(w') \geq d]$ if and only if $\exists d_{\text{nec}}$ such that $d_{\text{nec}} = \text{MAX}\{d : \forall w' \in \textsc{Acc}(w)[Q(x)(w') \rightarrow \textsc{adj}(x)(w') \geq d]\}$.

[7] A precise representation, both for exercisable capacities $\textsc{adj}_{\textsc{cap}}$, and for their manifestations, $\textsc{do-adj}_{\textsc{cap}}(x)(d)(w)$ is left for future work. Roughly, we would like a manifestation to be an event e such that $\textsc{adj}(x)(d)$ holds over the runtime $\tau(e)$ of e, and such that e does not satisfy the subinterval property.

[8] Schwarzschild (2008) also offers a paraphrase of *too* constructions that links the matrix adjective to the possibility/impossibility of the E&T complement as a reason or cause.

(22a) is odd because it suggests that the ring's price had an effect on Morgan's intelligence. In other words, the *because*-clause is interpreted as modifying an existing causal chain between intelligence and ring-buying. (22b), on the other hand, is fine. The fridge's emptiness affects its weight, and this impacts the causal connection between strength and fridge-lifting.[9]

As a final point in support of the causal analysis of E&T constructions involving exercisable capacities, consider the structure of the course of events leading from being fast to winning a race. If this involves an initial process (e.g. running) that culminates in the causal consequence (winning) – that is, if the event structure underlying (2) resembles an *accomplishment* – the implicative inference is supported. If the context supports, instead, a temporal separation between being fast and the conclusion of the race, the accomplishment structure will be broken. In this case, we predict implicative entailments to go away, although implicatures may still arise. Marques (2012) provides data from Portugese that supports this idea:

(23) *No último encontro, ele foi humilhado o suficiente para agora recusar o convite para um novo encontro (mas parece que já se esqueceu, porque está a pensar aceitar.)*
'In the last meeting, he was-PFV humiliated enough to now refuse the invitation for a new meeting (but it appears that he already forgot, since he is planning to accept).'

The requirement of an accomplishment structure supports the idea that a causal chain is involved in producing E&T entailments.

3.2 Implicativity and aspectual coercion

Proposal (20) holds that E&T constructions entail their complements just in case they entail manifestations of d_{nec}-ADJ. Perfective aspect ensures the desired manifestation.

(24) a. *Juno a été assez rapide pour gagner la course.* ⊢ *Juno won the race.*
'Juno was-PFV fast enough to win the race.'
b. *Juno n'a été pas assez rapide pour gagner la course.* ⊢ *Juno did not win the race.*
'Juno was-PFV not fast enough to win the race.'

(25) a. *Juno était assez rapide pour gagner la course.* ⊬ Juno won the race.
'Juno was-IMPF fast enough to win the race.'
b. *Juno n'était pas assez rapide pour gagner la course.* ⊬ Juno did not win the race.
'Juno was-IMPF not fast enough to win the race.'

Before combining with tense and aspect, $\text{ADJ}(x)(d)$ represents a stative predicate (of events).[10] The perfective, however, selects for eventive predicates (Dowty, 1986), providing existential closure over an event which is bounded within a reference time.

(26) $[\![\text{PFV}]\!] := \lambda R_v \lambda t.\exists e[R(e)\ \&\ \tau(e) \subseteq t]$

Statives can only combine with perfective aspect via *aspectual coercion*, which maps them to a compatible predicate type (Moens & Steedman, 1988; de Swart, 1998). (27) is an example of *inchoative* coercion, which maps the stative predicate to its initiation.

(27) *Jupiter a aimé Europa.* → *Jupiter fell in love with Europa.*
'Jupiter loved-PFV Europa.'

[9] Note, also, that replacing *empty* in (22b) with *full* again results in oddness, in this case because full things are typically assumed to be heavier than less full ones. This should not work in Marie's favour.

[10] That is, it satisfies the subinterval property: whenever $\text{ADJ}(x)(d)$ holds of an eventuality e, it also holds of every subeventuality $e' \sqsubseteq e$.

Homer (2011) identifies another type of coercion, which he calls *actualistic*. It takes a stative and returns a pragmatically-determined eventive predicate which coincides temporally with an event in the denotation of the original stative.[11]

(28) *Jean a eu du tact.* = PST(PFV(ACT(Jean have tact))) → *Jean acted tactfully.*
'Jean had-PFV tact.'

Predicates involving exercisable capacities are prime candidates for actualistic coercion. Unlike individual-level properties like *be tall*, *have tact* and *be fast* are necessarily associated with actions: the encoded capacity is the capacity to perform an action characterized by the matrix adjective. ACT maps $\text{ADJ}(x)(d)$ to such an action.

The combination of PFV and ACT in (24a) entails that Juno did something with speed d_{nec} within the reference time. This activates the sufficiency condition (20.III), deriving complement entailment. (24b) entails a manifestation of speed strictly less than d_{nec}, precluding complement realization. Thus, actualistic coercion ensures not only the implicative entailment, but also requires entails that Juno actually did something, even in the negative case. This is as desired: like negative implicative assertions, negative E&T are associated with the inference that some attempt was made to achieve the E&T complement.[12]

Imperfective aspect locates reference time within the situation time of an eventuality. It selects for statives. Assuming no other operators, the combination of past and imperfective in (25a) locates the reference time within the a situation of Juno having the capacity to do things at speed d_{nec}. No action is entailed, and no implicative inference arises.

(29) $[\![\text{IMPF}]\!] := \lambda R_v \lambda t. \exists e[R(e)\ \&\ \tau(e) \supset t]$

This is not the only possible use of imperfective aspect: it can also represent the progressive (PROG) or habitual/generic (HAB/GEN). The predictions change in these cases. PROG selects for eventive predicates (processes). Actualistic coercion applies, returning an in-progress event in which Juno manifests speed d_{nec}. The absence of an implicative entailment here follows from the accomplishment structure of E&T constructions noted in the preceding section. This is simply an instance of the *imperfective paradox*: in general, progressive descriptions of accomplishments can be true without entailing their results (*baking a cake* does not entail that a cake was baked; Dowty, 1979). I refer to Hacquard (2005) for the generic case.[13]

These points carry over to the negative imperfective case (25b). If we cancel the negative entailment by claiming the truth of a punctual event of race-winning, we are forced to reinterpret the matrix assertion either progressively or generically. In the former case, the imperfective paradox applies, this time to an ongoing manifestation of speed less than d_{nec}. In the latter, we necessarily interpret the race-winning event as abnormal – Juno is ordinarily not fast enough, but something unusual occurred in the case at hand.

What about English? It turns out that past-tense attributions of exercisable capacities are systematically ambiguous between eventive and stative readings:

(30) Juno was fast. → *eventive:* Juno did (something) fast/quickly.
→ *stative:* Juno had the capacity do (something) fast/quickly.

[11] Homer (2011) introduces actualistic coercion to derive *actuality entailments* (Bhatt, 1999), from perfectively-marked ability modals to the realization of their complements. I believe that ACT is more constrained in its output than Homer suggests, but will leave this discussion for future work.

[12] Actualistic coercion seems to be the default for exercisable capacity attributions under perfective marking, but it is not the only possibility. The introduction of an adverbial modifier like *suddenly/soudain* can privilege an inchoative interpretation, which would not entail a manifestation: thus, *Juno a soudain été assez rapide pour gagner la course* is predicted not to entail that Juno won the race.

[13] Under the current proposal, nothing mandates that Juno exercise speed d_{nec}, even in those normal worlds where she has the capacity.

Contexts which privilege the eventive will (defeasibly) activate the E&T sufficiency condition, leading to implicative inferences. This cannot be an entailment, because English never *entails* the eventive reading. Overt cancellation of the complement proposition will force a stative, progressive, or generic interpretation of the main clause. This provides some evidence for the reality of covert aspectual operators in languages like English.

(31) Juno was fast enough to win the race, but ... *stative:* ...she did not race.
progressive: ...she suddenly twisted her ankle and had to stop.
generic: ...unexpectedly, she did not run at her full speed.

3.3 Managing and doing

Finally, proposal (20) provides an explanation for the difference between implicatives and E&T constructions with respect to implicative entailment in the imperfective. The key point, again, is that E&T assertions are not inherently eventive. Implicative assertions are.

(3a) *Juno était assez rapide pour gagner la course, mais elle n'a jamais gagné.*
'Juno was-IMPF fast enough to win the race, but she never won.'

(4a) *Juno réussissait à gagner la course, #mais elle n'a jamais gagné.*
'Juno manage-IMPF to win the race, #but she never won.'

An implicative directly asserts the truth of the causing event. Unlike E&T constructions, this is not immediately compatible with imperfective aspect: the main clause of (4a), for instance, must produce a generic or habitual interpretation. Roughly, this consists of regular instances – across worlds or times – in which the causing event occurs in situations where it is both necessary and sufficient for the complement event.

This means that in all normal situations (or worlds), Juno wins the race. We predict this to be compatible with the denial of any one particular instance of race-winning, but not with the claim that she *never* won, as in (4a). (3a) is acceptable, on the other hand, because we only require that Juno have the capacity for speed d_{nec} in all normal situations where manifesting this speed is both necessary and sufficient for winning the race. There need not be any such situation in which she actually acted on this capacity.[14]

4 Conclusion

This paper proposes to explain Hacquard (2005)'s data on implicative entailments from perfectively-marked E&T constructions in terms of an underlying causal structure shared with implicative verbs. Independently-motivated aspectual contrasts on certain types of property attributions are recruited to explain the 'optional' quality of E&T implicativity. Specifically, where implicative assertions always assert the truth of a causing event, E&T constructions minimally assert the possibility of such an event. When this possibility is realized, the implicative inference follows, but not otherwise; this accounts for the difference between implicative and E&T assertions in the imperfective as well as in aspectually-unmarked languages.

A number of questions remain for further investigation. For instance, how does the presupposition of causal sufficiency arise when an E&T matrix predicate and complement are linked by causal necessity? Why is there no parallel sufficiency presupposition when the E&T

[14] It is worth noting that, in English, *manage* is at best marginally compatible with the progressive. I am not certain if these facts are the same in French, but the account given here would predict this to be the case.

relationship is a deontic (or epistemic) one? The link between exercisable capacities, causes, and the accomplishment structure underlying E&T assertions suggests a broader link between causality and concepts of disposition and ability. Looking ahead, this suggests a role for an implicative-style semantic structure in the longstanding puzzle of actuality entailments (Bhatt, 1999) from perfectively-marked ability modals to the realization of their complements.

References

Baglini, R. & Francez, I. 2016. The implications of managing. *Journal of Semantics* 33, 541–560.

Bhatt, R. 1999. Ability modals and their actuality entailments. In K. Shahin, S. Blake & E.-S. Kim (eds.), *Proceedings of the West Coast Conference on Formal Linguistics*, vol 17, pp. 74–87, Stanford: CSLI.

Bierwisch, M. 1987. Semantik der Graduierung. In M. Bierwisch & E. Lang (eds.), *Grammatische und konzeptuelle Aspekt von Dimensionsadjektiven*, vol 16 of *Studia Grammatica*, pp. 91–286, Akademie-Verlag.

de Swart, H. 1998. Aspect shift and coercion. *Natural Language and Linguistic Theory* 16, 347–385.

Dowty, D. 1979. *Word Meaning and Montague Grammar: The Semantics of Verbs and Times in Generative Semantics and in Montague's PTQ*. Dordrecht: Reidel.

Dowty, D. 1986. The effects of aspectual class on the temporal structure of discourse: semantics or pragmatics. *Linguistics and Philosophy* 9, 37–61.

Hacquard, V. 2005. Aspects of *too* and *enough* constructions. In *Proceedings of Semantics and Linguistic Theory*, vol 15.

Homer, V. 2011. French modals and perfective. In M. Washburn, K. McKinney-Bock, E. Varis, A. Sawyer & B. Tomaszewicz (eds.), *Proceedings of the West Coast Conference on Formal Linguistics*, vol 28, pp. 106–114, Somerville, MA: Cascadilla Press.

Karttunen, L. 1971. Implicative verbs. *Language* 47, 340–358.

Karttunen, L. & Peters, S. 1979. Conventional implicature. In Oh & Dinnen (eds.), *Syntax and Semantics*, pp. 1–56, New York: Academic Press.

Kratzer, A. 1981. The notional category of modality. In H.-J. Eikmeyer & H. Rieser (eds.), *Words, Worlds, and Contexts: New Approaches in Word Semantics*, de Gruyter.

Marques, R. 2012. Covert modals and (non-)implicative readings of *too/enough* constructions. In W. Abraham & E. Leiss (eds.), *Covert Patterns of Modality*, pp. 238–266, Cambridge Scholars Publishing.

Meier, C. 2003. The meaning of *too*, *enough* and *so ... that*. *Natural Language Semantics* 11, 69–107.

Moens, M. & Steedman, M. 1988. Temporal ontology and temporal reference. *Computational Linguistics* 14, 15–28.

Nadathur, P. 2016. Causal necessity and sufficiency in implicativity. In M. Moroney & J. Collard (eds.), *Proceedings of Semantics and Linguistic Theory 26*, pp. 1002–1021.

Schubert, L. & Pelletier, F. 1989. Generically speaking, or, using discourse representation theory to interpret generics. In G. Chierchia, B. Partee & R. Turner (eds.), *Properties, Types and Meaning*, vol II, Springer.

Schulz, K. 2011. If you'd wiggled A, then B would've changed. *Synthese* 179, 239–251.

Schwarzschild, R. 2008. The semantics of comparatives and other degree constructions. *Language and Linguistics Compass* 2.

von Stechow, A., Krasikova, S. & Penka, D. 2004. The meaning of German *um zu*: necessary condition and *enough/too*. Workshop on Modal Verbs and Modality, handout.

Turkish plural nouns are number-neutral: Experimental data

Agata Renans[1], George Tsoulas[2], Raffaella Folli[1], Nihan Ketrez[3], Lyn Tieu[4], Hanna de Vries[2], and Jacopo Romoli[1]

[1] Ulster University, Belfast, United Kingdom
am.renans@ulster.ac.uk, r.folli@ulster.ac.uk, j.romoli@ulster.ac.uk
[2] University of York, York, United Kingdom
george.tsoulas@york.ac.edu, hanna.devries@york.ac.uk
[3] Istanbul Bilgi University, Istanbul, Turkey
nihan.ketrez@bilgi.edu.tr
[4] Western Sydney University & Macquarie University, Sydney, Australia
lyn.tieu@gmail.com

Abstract

Across languages, plural marking on a noun typically conveys that there is more than one entity in the denotation of the noun. In English, this 'more than one' meaning is generally regarded as an implicature on top of a 'semantically unmarked'/number-neutral literal meaning of the plural noun ([10, 18, 20]; see also [5, 12]). In Turkish, however, it is controversial whether plural nouns should be analysed as number-neutral or whether they should directly denote strict plurality [2, 19, 6]. This debate is important as it can shed light on the meanings number marking can have across languages, thereby constraining cross-linguistically adequate theories of the semantics of number. We tested Turkish-speaking adults and 4–6-year-old children on the interpretation of plurals in upward- and downward-entailing contexts, as compared to the 'not all' scalar inference of *bazı* 'some'. The results of our experiment support a theory of plural nouns which includes a number-neutral interpretation.

1 Introduction

Across languages, plural marking conveys a multiplicity inference (MI): (1) is typically interpreted as giving rise to the interpretation that Tiger planted more than one tree. The same sentence with the corresponding singular noun in (2) does not give rise to this interpretation.

(1) Tiger planted trees.
$\rightsquigarrow$ *Tiger planted more than one tree* MULTIPLICITY INFERENCE (MI)

(2) Tiger planted a tree.
$\not\rightsquigarrow$ *Tiger planted more than one tree*

In the case of English, the MI is generally claimed not to be encoded in the literal meaning of the plural noun (see, for example, [10, 18, 20]). That is, the plural in English would not encode strict plurality, as in (3), but rather number-neutrality, as in (4):

(3) $[\![\text{tree-s}]\!] = \{a \oplus c, b \oplus c, a \oplus b \oplus c\}$ STRICT PLURAL

(4) $[\![\text{tree-s}]\!] = \{a, b, c, a \oplus b, a \oplus c, b \oplus c, a \oplus b \oplus c\}$ NUMBER NEUTRAL

Under the number neutrality approach, it remains to be explained how the MI arises. There are two competing proposals in the literature: according to the first, the MI arises as an implicature

[18, 20], while on the second approach, the MI corresponds to one of the possible meanings of an ambiguous plural [5, 12]. In other words, the relevant theoretical options regarding the interpretation of the plural across languages are as follows: either the plural directly encodes multiplicity via strict plurality or it includes a number-neutral denotation. If the latter, then the MI is either an implicature or it arises through an ambiguity in the meaning of the plural.[1,2] The general consensus is that the possible interpretations of plural nouns in English include a number-neutral reading.

Compared to English, the nature of the plural in Turkish is more controversial, with existing arguments for strict plurality on the one hand [1, 2] and for number neutrality on the other [8, 19].[3] We conducted an experimental study in order to contribute to this debate. We tested Turkish-speaking adults and 4-6-year-old children on the interpretation of plural nouns in upward- and downward-entailing contexts and compared them to the 'not all' scalar inference of *bazı* 'some'. When combined with the assumption that Turkish plural nouns are not scopally inert [3, 2], the experimental results support a number-neutral theory of Turkish plural nouns. We argue that our results support an analysis of the Turkish plural that is very similar to what has been assumed for the English plural, and suggest that the source of observed differences can be found in different scope possibilities associated with bare plurals in the two languages.

The rest of the paper is structured as follows. Section 2 presents the semantics of plural nouns in Turkish. In particular, we discuss the two main approaches proposed in the literature: strict plurality and number neutrality. In Section 3, we discuss the predictions of the two approaches. We present our experimental study in Section 4, and discuss the results in the context of the theoretical predictions in Section 5. Section 6 concludes the paper.

2 Background: Two approaches to Turkish plural nouns

As in English, Turkish plural nouns give rise to a multiplicity inference, as in (5):

(5) Kaplan ağaç-lar ek-ti.
tiger tree-PL plant-PAST
'Tiger planted trees.' $\rightsquigarrow$ *Tiger planted more than one tree* MI

As mentioned, there is still controversy as to whether Turkish plural nouns should denote strict plurality directly or whether they should be associated with a number-neutral interpretation, with the MI arising as an implicature or through ambiguity.

The situation is complicated by the fact that bare plural nouns in Turkish, unlike in English, have been argued not to be scopally inert [3, 2]. That is, while in English a bare plural noun like *doctors* can only scope below *want* in (6) [4], it is claimed that in the Turkish counterpart of this sentence, *doktorlar* 'doctors' can take scope either above or below *want*, as in (7) [3, 2]:

(6) Mary wants to meet doctors. [from 3, p.51]
$\approx$ *Mary wants to meet some doctors or other.* *want* > *doctors* (narrow scope)

(7) Mary doktor-lar bul-mak ist-iyor.
Mary doctor-PL meet-INF want-PROG.3

[1] In other words, the plural is number-neutral under the implicature approach, while it is ambiguous between a number-neutral version and a strictly plural one in the ambiguity approach. Since under both approaches number-neutrality is among the possible denotations of the plural, we refer to both as *number-neutral* approaches.

[2] There is a third type of approach based on homogeneity [11]. We leave to future research a thorough evaluation of the homogeneity account against our data.

[3] [1, 2] actually focus on Western Armenian but claim that their analysis would extend to Turkish.

'Mary wants to meet doctors.' [from 3, p.51]
≈ *Mary wants to meet some doctors or other* *want* > *doctors* (narrow scope)
≈ *There are some doctors that Mary wants to meet* *doctors* > *want* (wide scope)

This suggests that a negative sentence containing a bare plural in Turkish, such as (8), might also in principle give rise to two interpretations depending on the scopal relation between the plural and negation.[4]

(8) Kaplan ağaç-lar ek-me-di.
tiger tree-PL plant-NEG-PAST
'Tiger didn't plant trees.'
a. ≈*It's not true that Tiger planted trees* (NEG > PL)
b. ≈*There are some trees that Tiger didn't plant* (PL > NEG)

3 Predictions

3.1 The effect of polarity

Both the strict plural and number-neutral approaches, combined with the assumption that bare plurals in Turkish can take wide scope over negation, make clear predictions regarding the acceptability of sentences with plural nouns in positive vs. negative contexts. Consider first a positive sentence such as the one in (9). Both the strict plurality and number neutrality predict (9) to be unacceptable in the described context in which Tiger planted only one tree.[5,6]

(9) **Context:** Tiger planted only one tree.
Kaplan ağaç-lar ek-ti.
tiger tree-PL plant-PAST
'Tiger planted trees.'
STRICT PLURALITY APPROACH: ≈ *Tiger planted more than one tree*
NUMBER NEUTRALITY APPROACH: ≈ *Tiger planted more than one tree*

However, the predictions of the two approaches diverge in the case of negation. Consider first the interpretation on which the plural scopes under negation. In this case, the meaning of (10) under both approaches can be paraphrased as in (10-a) and (10-b):

(10) **Context:** Tiger planted only one tree.
Kaplan ağaç-lar ek-me-di.
tiger tree-PL plant-NEG-PAST
'Tiger didn't plant trees.'
a. STRICT PLURALITY: ≈ *Tiger didn't plant more than one tree* (NEG>PL)
b. NUMBER NEUTRALITY: ≈ *Tiger didn't plant any tree* (NEG>PL)
c. BOTH APPROACHES:≈ *There are trees that Tiger didn't plant* (PL>NEG)

[4]Note however that the intuition that bare plurals can get a wide scope interpretation with respect to negation is not shared by all Turkish native speakers.

[5]As mentioned above, when we refer to the number neutrality approach, we mean an account that includes a number-neutral interpretation among the possible readings of the plural, combined with either an implicature or an ambiguity approach to explain the presence of the MI.

[6]The number-neutral approach actually predicts the two readings in both the positive and negative condition. One of the two reading is assumed to be preferred by a principle which favours the strongest reading in each case.

	STRICT PLURALITY	NUMBER NEUTRALITY
Tiger planted trees.	×	×
Tiger didn't plant trees. (NEG>PL)	✓	×
Tiger didn't plant trees. (PL>NEG)	✓	✓

Table 1: Predicted acceptability of positive and negative sentences in a context in which Tiger planted only one tree, according to the strict plural and number-neutral approaches.

As is clear from the paraphrases, while the strict plurality approach predicts the sentence in (10) to be acceptable as a true description of the described context (it is true that Tiger didn't plant more than one tree), the number-neutral approach predicts (10) to be judged false in the same context (it is false that Tiger didn't plant any trees). The situation differs, however, when the plural takes wide scope over negation. In this case, under both approaches, the sentence in (10) can be roughly paraphrased as in (10-c). Since this interpretation is compatible with the described context, both approaches predict (10) to be acceptable in this context.

To sum up, both approaches predict the positive sentence *Tiger planted trees* to be judged false in a context in which Tiger planted only one tree. However, while the strict plural approach predicts invariable acceptance of negative sentences in the same context, the number-neutral approach makes different predictions depending on the scopal interaction between the plural and negation. The predictions of the two approaches are summarized in Table 1.

3.2 Multiplicity inferences vs. standard implicatures

The number-neutral approach, combined with the view that the MI is a scalar implicature (SI), makes further predictions regarding the relationship between the MI and other SIs, and in particular in relation to whether children are expected to compute these inferences, compared to adults. Specifically, the implicature approach predicts that children should access the 'more than one' meaning less than adults do, mirroring the behavioral pattern associated with standard lexical-scale-based implicatures (for relevant discussion, see [21, 22, 17, 25]). The strict plurality approach, on the other hand, makes no particular predictions with respect to the comparison between the MI and other SIs, in children vs. adults.

4 Experiment

We tested the predictions of the strict plurality and number neutrality approaches discussed above by investigating Turkish speakers' interpretations of plural nouns in positive and negative contexts, as well as comparing the MI in Turkish to the SI of *bazi* ('some') [22, 17].

4.1 Methods

Participants We tested 45 adult native speakers of Turkish and 22 Turkish-speaking children (age range 4-6 years, mean age 5;02). One child and three adults were excluded from the analysis for failing to pass the control trials, leaving a total of 42 adults and 21 children.

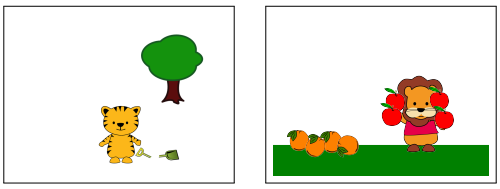


Figure 1: *Left:* image for the MI targets in (11)/(11); *Right:* image for the SI target in (12).

Procedure Participants listened to short stories presented through a PowerPoint presentation on a laptop computer. Participants were introduced to a puppet with whom they interacted throughout the experiment via webcam (in reality, through pre-recorded videoclips). After each story, the experimenter asked a question to the puppet and the puppet responded with the test sentence. The participants' task was to judge the puppet's utterances by rewarding her with one, two, or three strawberries, depending on her performance [9, 23]. Participants were instructed to give the puppet one strawberry if they thought the puppet didn't answer well, three strawberries if she answered well, and two strawberries if the puppet's answer was somewhere in the middle – not perfect, but somewhat okay.

Materials We manipulated three factors: Group (child vs. adult), Inference Type (MI vs. SI), and Polarity within the plural condition (positive vs. negative). The MI and SI conditions were introduced in blocks and their order was counterbalanced across participants. To illustrate, consider first an example of a positive and a negative MI target, in (11). The story made it clear that the MI triggered by the plural was not satisfied in the context. The corresponding picture is provided in Figure 1.[7]

(11) **Context:** Tiger only planted this one tree and no flowers. MI target
EXP: Okay, Ellie, so Tiger didn't plant any flowers. What about trees?

a. Kaplan ağaç-lar ek-ti.
tiger tree-PL plant-PAST
'Tiger planted trees.'

b. Kaplan ağaç-lar ek-me-di.
tiger tree-PL plant-NEG-PAST
'Tiger didn't plant trees.'

On both the strict plurality and number neutrality approaches, participants were expected to interpret the positive targets with the MI (i.e. *Tiger planted more than one tree*); given that the MI was not satisfied in the context, participants were expected to reward the puppet with either one or two strawberries. Note that the number-neutral approach allows for a number-neutral reading without the MI, while the strict plural approach does not; the former, but not the latter, would therefore allow for variation in participants' responses to the positive targets.

Moving on to the negative targets, the strict plurality approach predicts that participants should invariably access the strict plural interpretation of the noun. Crucially, this interpretation is compatible with the given context, irrespective of the scopal interaction of the plural and negation. Therefore, participants were expected to give the puppet the maximal reward. As for the number-neutral approach, the expected reward would depend on the scope of the plural: the interpretation on which negation scopes over the plural is incompatible with the context, which would lead to non-maximal rewards in contexts like that in Figure 1. The interpretation on which the plural scopes over negation, on the other hand, is compatible with the context

[7] To keep things interesting for the child participants, the characters and objects varied from one item to the next.

and participants were therefore expected to select the maximal reward.

In the SI condition, borrowed from [21, 22], it was made clear in the stories that the action of the protagonist involved the whole set of pictured objects. When asked what had happened in the story, the puppet answered using the scalar term *bazı* 'some', as in (12) (see Figure 1 for the corresponding picture):

(12) **Context:** Lion took no oranges and all of the apples. SI target
EXP: Okay, Ellie, so the Lion didn't carry any oranges. What about the apples?
PUPP: Aslan elma-lar-ın bazı-lar-ı-nı taşı-dı.
Lion apple-PL-GEN some-PL-POSS.3SG-ACC carry-PAST
'Lion carried some of the apples.'

If participants computed the SI of *bazi* 'some', i.e. *the lion didn't carry all of the apples*, they were expected to reward the puppet with one or two strawberries. If the utterance instead was interpreted literally, participants were expected to give the puppet the maximal reward.

Participants also received eight control trials to ensure that they could give minimal and maximal rewards where appropriate. Four of the controls corresponded to clearly true plural sentences that were expected to elicit the maximal reward, as in (13) and (14):

(13) **Context:** Giraffe did not bake any cakes but she baked four cookies.
EXP: Okay, Ellie, so Giraffe didn't bake any cakes. What about cookies?
PUP:Zürafa kurabiye-ler pişir-di.
Giraffe cookie-PL cook-PAST
'Giraffe baked cookies.' Positive control

(14) **Context:** Sheep baked four pizzas but no baklavas.
EXP:Okay, Ellie, so Sheep baked pizzas. What about baklavas?
PUP:Koyun baklava-lar pişir-me-di
Sheep baklava-PL cook-NEG-PAST
'Sheep didn't bake baklavas.' Negative control

Four other controls corresponded to clearly true or clearly false negative sentences that contained a definite noun phrase instead of a bare plural, which allowed us to check that participants could correctly interpret negation independently of the plural. These trials could be associated with either a minimal or a maximal reward target; the experimenter selected the appropriate version of the trial depending on how participants responded to the critical target trials, balancing the overall number of minimal and maximal rewards given across the experiment.

(15) **Context:** Zebra painted four vases and no bowls. Negation control
EXP:Ellie, can you tell us something about the story?

PUP': Zebra kase-ler-i boya-ma-dı!
Zebra bowl-PL-ACC paint-NEG-PAST
'Zebra didn't paint the bowls!'

PUP":Zebra vazo-lar-ı boya-ma-dı!
Zebra vase-PL-ACC paint-NEG-PAST
'Zebra didn't paint the vases!'

In sum, each participant received two training items followed by 18 test trials: 6 critical plural targets (3 positive, 3 negative), 4 SI targets, 4 clearly true positive and negative plural controls, and 4 clearly true or clearly false negation controls. The MI and SI targets were presented in blocks, counterbalanced across participants, and the test and control trials within the plural block were pseudo-randomized.

4.2 Results

Figure 2 displays the proportion of 1-, 2-, and 3-strawberry responses to the positive, negative, and scalar implicatures targets. As a first step, we collapsed the 'non-maximal' 1- and 2-strawberry responses. For the SI and positive MI targets, 1- and 2-strawberry responses were interpreted as a measure of the target inference having been computed, while 3-strawberry responses corresponded to a reading without the relevant inference. For the negative plural targets, 3-strawberry rewards were interpreted as consistent with the MI having been computed (under negation), while 1- and 2-strawberries corresponded to a reading without the MI.

Focusing on the plural positive targets, we observe that adults mostly rejected the positive sentences in contexts that were incompatible with the MI, indicating they had computed the inference. By contrast, children tended to accept the positive sentences in the same contexts, suggesting they hadn't computed the MI. On the plural negative targets, on the other hand, adults appeared to split between selecting the maximal and the non-maximal rewards. Children instead tended to give minimal rewards only, suggesting that they interpreted the plural under negation without the MI. As for the SI condition, both groups generally selected non-maximal rewards, indicating they computed the implicature of *bazı* 'some'.

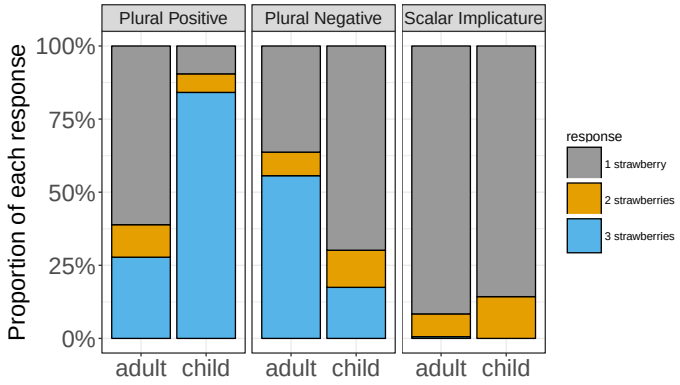


Figure 2: Proportion of 1-, 2-, and 3-strawberry responses across conditions.

Figure 3 presents the results for the positive and negative plural targets, with the ternary responses recoded in binary terms. Logistic regression models fitted to these recoded plural data revealed a significant effect of Group ($\chi(2) = 29.2$, $p < .001$) but no effect of Polarity or Group:Polarity interaction. Finally, in Figure 3 we also compare the positive MI targets with the SI targets, across the two groups (with ternary responses recoded in binary terms). Both groups generally computed more implicatures than 'more than one' meanings, with adults computing more multiplicity inferences than children ($\chi(2) = 29.1$, $p < .001$).

To sum up, the results indicate that while adults mostly computed MIs from positive plural sentences, they were split in their interpretation of plural negative sentences. Children, on the other hand, did not interpret the plural sentences as giving rise to MIs in either condition. Finally, both adults and children computed the scalar implicature of *bazı* 'some'.

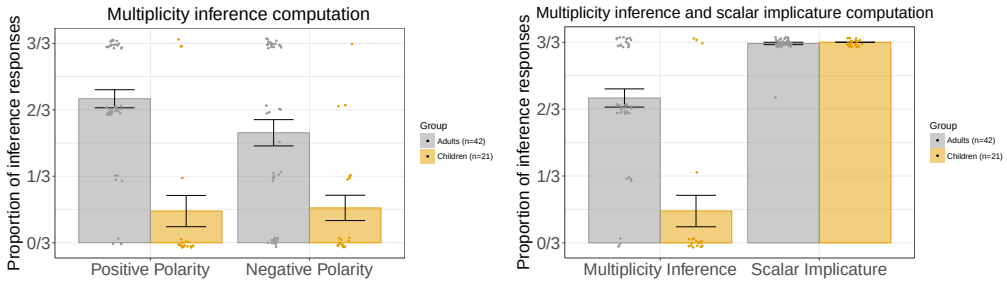


Figure 3: Multiplicity inference computation in positive and negative contexts (left) and multiplicity and scalar inference computation (right) in positive contexts, after recoding the ternary responses in binary terms (1 for inference reading, 0 for no inference reading). Each dot represents an individual participant's mean inference rate for the given condition.

5 Discussion

Adult data As predicted by both the strict plurality and number neutrality approaches, adults mostly rejected positive plural sentences in contexts in which the 'more than one' meaning was falsified, replicating the English and Greek findings reported in [22] and [17]. Moreover, while the variance in adults' responses pose a challenge for the strict plurality approach, it is nevertheless in line with the number-neutral approach.

As for negative plural sentences, adult participants were split between acceptance and rejection of the negative targets, as predicted by the number neutrality approach but not by the strict plurality approach. The split we observe among participants could be due to their accessing different scopal interpretations of the plural with respect to negation (PL>NEG or PL<NEG), as discussed in Section 3. In this respect, the adults' data differed from those reported for English [22] and Greek [17], where participants tended to overwhelmingly reject the negative target sentences. Put differently, the results of similar experiments in English and Greek reveal a much larger proportion of number-neutral interpretations under negation. This difference, we argue, could be due to the differences in scopal possibilities in English and Greek compared to Turkish: while in Turkish the plural is able to scope either below or above negation, in English and Greek only the former option is available.

Child data The results from the positive targets are consistent with the implicature approach: children showed a strong preference for the number-neutral interpretation of plural nouns, in contrast to adults. In particular, the child participants interpreted the positive targets on their number-neutral meaning, in line with the observation that they tend to derive fewer implicatures at this age [15, 16, a.o.]. On the negative targets, on the other hand, children, unlike adults, failed to access a wide scope interpretation of the plural above negation, which is also in line with some existing literature that indicates children in this age range have a preference for isomorphic interpretations of scopally ambiguous sentences [14, 13, 7]).

While the overall results are in line with the implicature approach, the results in the SI condition are surprising for this approach. In particular, participants did not treat the MI and SI targets alike: children generally did not compute MIs but they always computed the SI of *bazı* 'some', like adults. While variation among scalar terms has been observed in adults [24],

such an extreme difference between the MI and SI conditions warrants further investigation.[8]

6 Conclusion and future directions

The present study examined whether plural nouns in Turkish denote strict plurality or are number-neutral given the existence of competing view in the theoretical literaure on Turkish. We tested adults and 4-6-year-old children on the interpretation of bare plural nouns in upward- and downward-entailing contexts, and compared this to the scalar implicature of *bazi* 'some'. Overall, when combined with the assumption that bare plurals in Turkish are not scopally inert [3, 2], our results support a theory of Turkish plurals as number-neutral. Moreover, both the adults' and children's results are in line with an implicature approach. However, the results from the scalar implicature condition are challenging for this approach, even when we consider that there is reported variability among scalar terms in adults [24]. Future work could test whether scope is indeed a factor in the interpretation of Turkish plural nouns, by directly manipulating whether the potential wide-scope reading of the bare plural is made true in the context.

References

[1] Alan Bale, Michaël Gagnon, and Hrayr Khanjian. Cross-linguistic representations of numerals and number marking. In Nan Li and David Lutz, editors, *Semantics and Linguistic Theory (SALT) 20*, pages 582–598, 2010.

[2] Alan Bale and Hrayr Khanjian. Syntactic complexity and competition: The singular-plural distinction in western armenian. *Linguistic Inquiry*, 45(1):1–26, 2014.

[3] Heather Bliss. The semantics of bare nuoun in Turkish. *Calgary Papers in Linguistics*, 25(1):1–65, 2004.

[4] Gregory N. Carlson. *Reference to Kinds in English.* PhD thesis, University of Massachusetts, Amherst, 1977.

[5] Donka Farkas and Henriette de Swart. The semantics and pragmatics of plurals. *Semantics & Pragmatics*, 3(6):1–54, 2010.

[6] Emrah Görgülü. *Semantics of nouns and the specification of numner in Turkish.* PhD thesis, Simon Fraser University, 2012.

[7] Andrea Gualmini, Stephen Crain, Luisa Meroni, Gennaro Chierchia, and Maria T. Guasti. At the semantics/pragmatics interface in child language. In R. Hastings, B. Jackson, and Z. Zvolenszky, editors, *Semantics and Linguistic Theory (SALT) 11*, pages 231–247, 2001.

[8] Seda Kan. Number marking and Turkish Noun Phrases. ms., Universiy of Massachusetts, Amherst, 2010.

[9] Napoleon Katsos and Dorothy V.M. Bishop. Pragmatic tolerance: Implications for the acquisition of informativeness and implicature. *Cognition*, 120(1):67–81, 2011.

[8] The calculation of SI by children could also be influenced by the genitive partitive construction used in the SI targets. We will investigate it further in future research.

[10] Manfred Krifka. Nominal reference, temporal constitution and quantification in event semantics. In Renate Bartsch, editor, *Semantics and contextual expressions*, pages 75–116, 1989.

[11] Manuel Križ. Bare plurals, mutlitplicity, and homogeneity. Ms., Institut Jean Nicod, 2017.

[12] Luisa Martí. Numerals and the the theory of number. Msc. Queen Mary, University of London, 2017.

[13] J. Musolino and Lidz J. The scope of isomorphism: turning adults into children. *Language Acquisition*, 11(4):277–291, 2003.

[14] Julien Musolino. *Universal grammar and the acquisiton of semantic knowledge: an experimental investigation into the Acquisition of quantifier-negation interaction in English.* PhD thesis, University of Maryland, 1998.

[15] Ira Noveck. When children are more logical than adults: experimental investigations of scalar implicatures. *Cognition*, 78(8):165–188, 2001.

[16] Anna Papafragou and Julien Musolino. Scalar implicatures: experiments at the semantics–pragmatics interface. *Cognition*, 86(3):253–282, 2003.

[17] Agata Renans, Jacopo Romoli, Maria-Margarita Makri, Lyn Tieu, Hanna de Vries, Raffaella Folli, and George Tsoulas. Abundance infernce of pluralised mass nouns is an implicature: Evidence from Greek. Ms. Ulster University and University of York, 2017.

[18] Uli Sauerland, Jan Anderssen, and Kazuko Yatsushiro. The plural is semantically unmarked. In S. Kesper and M. Reis, editors, *Linguistic evidence*, pages 413–434. Mouton de Gruyter, 2005.

[19] Yağmur Sağ. The semantics of numeral constructions in Turkish. In *Sinn und Bedeutung (SuB) 22*, 2017.

[20] Benjamin Spector. Aspects of the pragmatics of plural morphology: On higher-order implicatures. In Uli Sauerland and Penka Stateva, editors, *Presupposition and Implicature in Compositional Semantics*. Palgrave, 2007.

[21] Lyn Tieu, Cory Bill, Jacopo Romoli, and Stephen Crain. Plurality inferences are scalar implicatures: Evidence from acquisition. In Todd Snider, editor, *Proceedings of the 24th Semantics and Linguistic Theory Conference*, pages 122–136, 2014.

[22] Lyn Tieu, Cory Bill, Jacopo Romoli, and Stephen Crain. Plurality inferences are scalar implicatures: evidence from acquisition. Ms., Marquarie University and Ulster University, July 2017.

[23] Lyn Tieu, Manuel Križ, and Emmanuel Chemla. Children's acquisition of homogeneity in plural definite descriptions. Children's acquisition of homogeneity in plural definite descriptions, 2017.

[24] Bob van Tiel, Emiel van Miltenburg, Natalia Zevakhina, and Bart Geurts. Scalar diversity. *Journal of Semantics*, doi:10.1093/jos/ffu017, 2014.

[25] Kazuko Yatsushiro, Uli Sauerland, and Artemis Alexiadou. The unmaunmarked of plural: Crosslinguistic data. In Maria LaMendola and Jennifer Scott, editors, *41st annual Boston University Conference on Language Development*, pages 153–765, 2017.

Tense and Mood in Counterfactual Conditionals: The View from Spanish*

Maribel Romero

University of Konstanz, Konstanz, Germany
maribel.romero@uni-konstanz.de

Abstract

Counterfactual conditionals in Spanish are expressed using an additional layer of past tense –'fake' tense– and subjunctive mood. An analysis is developed in which each of these two pieces of morphology receives a uniform semantics independently motivated across the grammar: 'Fake' tense is analysed within the temporal remoteness line and subjunctive follows [Sch05]'s analysis of mood in complement clauses.

1 Introduction

The Spanish conditional sentences (3)-(4) give rise to the (defeasible) inference of counterfactuality, just like their respective English translations (1)-(2) [Lew73, And51]. Leaving Severe Tense Mismatch cases aside [Ipp03, Ipp13], we will refer to these structures as Counterfactual Conditionals (CCs). (At least) two pieces of verbal morphology are essential to produce a grammatical CC structure in Spanish and other Romance languages. First, like English, an additional layer of past tense –known as 'fake' tense– is needed. That is, even though the *if*-clause in (3) is concerned with an event happening at present time, that event is described using Past tense (i.e., 'had'); and even though the event in the *if*-clause in (4) is understood as directly preceding the speech time, it is described using Past Perfect (i.e., 'had gone'). Second, unlike English, where there is no (productive) mood distinction between indicative and subjunctive, the antecedent clause has to appear in the subjunctive.

(1) If Juan had a hang-over (right now/today), he would be in bed. PRESENT

(2) If Juan had gone to the party yesterday, the party would have been fun. PAST

(3) Si Juan tuviese resaca (ahora/hoy), (pro) estaría en la cama. PRESENT
If Juan **had.**SUBJ hang-over (now/today), (he) **would-be** in the bed

(4) Si Juan hubiese ido a la fiesta (ayer), la fiesta habría sido
If Juan **had.**SUBJ **gone** to the party (yesterday), the party **would.have been**
divertida. PAST
amusing.

If either of these two ingredients is removed, the sentences are not CC anymore. Keeping Subjunctive mood but removing the additional tense layer leads to ungrammaticality in the case of (3) and to a hypothetical interpretation in the case of (4).[1] Keeping the additional tense

*We thank two anonymous reviewers for their helpful pointers. The work leading to this paper was supported by the Research Unit 1614"What if?" funded by the Deutsche Forschungsgemeinschaft (DFG).

[1]Removing one layer of past tense in the *if*-clause and matrix clause of (4) results in (i), which can be used in a situation where time travel is possible and, hence, it is possible today for John to go to yesterday's party.

(i) Si Juan fuese a la fiesta (ayer), la fiesta sería divertida.
If Juan **went.**SUBJ to the party (yesterday), the party **would.be** amusing.

layer and removing Subjunctive mood leads, as will see, to structures that are grammatical in certain linguistic environments but that, again, have no counterfactual interpretation.

The present paper exploratorily develops an analysis of Spanish CCs that assigns each of these two pieces of morphology a uniform semantics independently motivated across the grammar. On the one hand, the additional past tense will not be interpreted modally, as has been done in the so-called modal remoteness approach to 'fake' tense in English ([Iat00, Sch14]). Instead, building on the temporal remoteness account by [Dud84, GvS09] a.o. and modifying [Rom14], it will interpreted temporally, as independently needed for Sequence of Tense. On the other hand, mood morphology will be treated as imposing a restriction on the world pronoun, as independently argued for complement clauses in Romance [Sch05]. In the bigger picture, the present paper contributes one more step in the formalization of the temporal remoteness approach to 'fake' tense, bringing it closer to a stage in which it can be compared to fully worked out analyses in the modal remoteness line like [Sch14]'s.

The paper is organized as follows. Section 2 presents [vS09]'s account of Sequence of Tense and applies it to complement clauses containing future indicative conditionals. Section 3 presents [Sch05]'s approach to mood in Romance complement clauses. Section 4 combines these two ingredients and presents the proposal. Section 5 concludes.

2 Additional past

It is known that an additional layer of past tense morphology is used in past attitude reports in indirect speech, a phenomenon known as 'Sequence of Tense' (SoT) [Abu97, Kus05, vS09]. Some simple examples are in (5)-(7). An event described using (Simple) Present in direct speech is described using (Simple) Past in indirect speech, as in (5). Similarly, an event described using Present Perfect in direct speech is described using Past Perfect in indirect speech, as in (6). Finally, an event described using a Future form in direct speech is described using the (corresponding) Conditional form in indirect speech, as in (7).[2] The same pattern obtains in Spanish and other Romance languages.

(5) a. Annalea said (last week): "Lucía **is** sick".
b. Annalea said (last week) that Lucía **was** sick.

(6) a. Annalea said (last week): "Lucía **has arrived** on time".
b. Annalea said (last week) that Lucía **had arrived** on time.

(7) a. Annalea said (last week): "Lucía **will come**".
b. Annalea said (last week) that Lucía **would come**.

To make the similarity between SoT and our Spanish CCs (3)-(4) more apparent, let us see how certain conditional structures change their verbal morphology when transferred from direct to indirect speech. Consider first scenario (8), which describes in direct speech Ana's past thought about a certain date –a salient temporal *res*, e.g., today December 20, 2017. The content of the thought is a future indicative conditional concerning hypothetical events *on* that date. When we describe this thought in direct speech, we have (9). When we describe this thought in indirect speech, temporal morphology shifts as indicated above: Present in the *if*-clause turns to Past and Future in the consequent clause turns to Conditional, as shown in (10). Interestingly, our CC (3) and the complement clause in (10) have exactly the same tenses and differ solely in the mood of the antecedent clause.

[2]Historically, *would* is the past form of *will* in English and the Spanish Future and Conditional arise from the present and past forms respectively of the same verbal periphrasis.

(8) Scenario: Ana was wondering in summer 2017 how things would be on Dec 20, 2017. She thought: "If Juan has a hang-over (that day), he will be in bed".

(9) Si Juan tiene resaca (ese día), (pro) estará en la cama.
If Juan **has**.IND hang-over (that day), he **will.be** in the bed

(10) Ella pensó que, si Juan tenía resaca, (pro) estaría en la cama.
She thought that, if Juan **had**.IND hang-over, he **would.be** in the bed
'She thought that, if Juan had a hang-over, he would be in bed.'

Consider now scenario (11), which describes in direct speech Ana's past thought of a future indicative conditional concerning hypothetical events *prior* to a certain date (e.g., today December 20, 2017). When the thought is expressed in direct speech, we have (12). When expressed in indirect speech, a parallel shift in temporal morphology obtains: Present Perfect in the *if*-clause turns to Past Perfect and Future Perfect in the consequent clause turns to Conditional Perfect, as shown (13).[3] Again, the tenses of our CC (4) and the complement clause in (13) are exactly the same, the two clauses differing only in the mood of the *if*-clause.

(11) Scenario: Ana was wondering in summer 2017 how things would be on Dec 20, 2017. She thought: "If Juan has gone to the party (the night before), the party will have been fun".

(12) Si Juan ha ido a la fiesta, la fiesta habrá sido divertida.
If Juan **has**.IND **gone** to the party, the party **will.have been** fun

(13) Ella pensó que, si Juan había ido a la fiesta, la fiesta habría sido
She thought that, if Juan **had**.IND **gone** to the party, the party **would.have been** divertida.
fun
'She thought that, if Juan had gone to the party, the party would have been fun.'

Let us see how these forms in indirect speech can be analysed under current theories of tense. We start with assumptions about LF structure. (Interpretable) tense morphology is treated like pronouns ([Par73] among many others). This means that, at LF, a temporal morpheme introduces a variable –represented as pro_i at LF– and some temporal feature –which we will write as a superscript at LF–, as illustrated in (14). Note that the superscripted temporal relation is relative to an anchor time variable pro_j. This is because, besides absolute uses of tense taking the utterance time t_0 as the anchor time, there are also relative uses taking other temporal variables as anchor [vS95, Abu97, Kus05]. Additionally, some pieces of temporal morphology may be left uninterpreted when licensed in a chain headed by an temporal pronoun with an interpretable PAST feature [GvS09, Rom14]. This is the case of the past tense layer in 'had.IND' and 'would.be' in e.g. our example (10), licensed by the c-commanding interpretable past tense in 'thought'. Such uninterpretable bits will appear crossed out in our LFs. Finally, the future indicative conditional is headed by a silent modal with a metaphysical modal base METAPHY and a stereotypical ordering source L (cf. [Kau05]). This leads to the LF (15) for our first indirect speech example (10):

(14) LF of past tense morpheme *-ed*: $\text{pro}_i^{[\text{PAST pro}_j]}$

[3] It is also possible to have the direct speech version (12) with Present Perfect ('has been') in the consequent clause (see [Kau05] on the difference between future indicative conditionals with and without *will*). The corresponding indirect speech version would be like (13) but with Past Perfect ('had been') in the consequent.

(15) LF: [λ0 Ana think at $\text{pro}_1^{[\text{PAST } \text{pro}_0]}$ [λ2 $\text{MODAL}^{\text{L}}_{\text{METAPHY}}$ pro_2
[λ3 ~~past~~ $[\text{pro}_4^{[\text{FUT } \text{pro}_3]}$ λ7[John have hang-over at pro_7]]]
[λ3 ~~past~~ $[\text{pro}_4^{[\text{FUT } \text{pro}_3]}$ λ7[John be in bed at pro_7]]]]]

Semantically, temporal features are interpreted as imposing presuppositions on the value of the variable [Hei94, Kra98], as defined in (16)-(18). We assume that the value of a temporal(/mood) variable, e.g. g(i), is an index, that is, a world-time pair. Temporal and accessibility constraints on indices are understood as in (19):

(16) $[\![past]\!]^g = [\![pro_i^{[\text{PAST } pro_j]}]\!]^g$ is defined only if g(i) < g(j);
if defined, $[\![pro_i^{[\text{PAST } pro_j]}]\!]$ = g(i)

(17) $[\![pres]\!]^g = [\![pro_i^{[\text{PRES } pro_j]}]\!]^g$ is defined only if g(i) $\circ$ g(j);
if defined, $[\![pro_i^{[\text{PRES } pro_j]}]\!]$ = g(i)

(18) $[\![fut]\!]^g = [\![pro_i^{[\text{FUT } pro_j]}]\!]^g$ is defined only if g(j) < g(i);
if defined, $[\![pro_i^{[\text{FUT } pro_j]}]\!]$ = g(i)

(19) a. For any two indices <w,t> and <w′,t′>:
<w,t> < <w′,t′> iff w=w′ and t is prior to t′.
<w,t> $\circ$ <w′,t′> iff w=w′ and t and t′ overlap.
b. For any two indices <w,t> and <w′,t′>:
<w,t> $\in$ MOD(<w′,t′>) iff t=t′ and w′ is accessible from w via MOD.

This gives us the semantic derivation in (20) for our example (10). After locally accommodating some of the temporal presuppositions, we obtain the truth conditions in (20c). Note that the pronoun pro_4 remains unbound and refers to an index i_4 whose temporal coordinate is a salient *res* time, namely, today December 20, 2017 in our scenarios.[4]

(20) a. Antecedent clause: λi_3: $i_3 < i_4$. John have hang-over at i_4
b. Consequent clause: λi_3: $i_3 < i_4$. John be in bed at i_4
c. Sentence: λi_0: $i_1 < i_0$. $\forall i_2 \in \text{Dox}_{\text{Ana}}(i_1)$ $\forall i_3 \in \text{Metaph}^{\text{L}}$ (i_2):
$i_3 < i_4 \wedge$ J have hang-over at i_4 $\rightarrow$ $i_3 < i_4 \wedge$ J be in bed at i_4

Our example (13) receives a parallel analysis. At LF, the c-commanding 'thought' licences one layer of uninterpreted past tense in the antecedent and consequent clauses, as before. But, since now we have Past Perfect and Conditional Perfect, we still have on layer of past tense to interpret in each clause, represented as $pro_6^{[\text{PAST } \text{pro}_5]}$ in the LF (21).[5] The semantic derivation proceeds a before, leading to the truth conditions in (22c):

[4]The treatment of pro_4 and i_4 in the text is a simplification. Since, in our scenario, Ana is having a *de re* thought about i_4, a proper treatment of it should include the acquaintance relation under which Ana accesses this *res*. This could be achieved by extending concept generators (CG) on *res* of e-type and more complex types ([PS03]) to *res* of s×t-type. The CG needed here would have to be as indicated in (i), where w(i) is the world-member of index i, t(i) is the time-member of i and α stands for the identifying property under which Ana is acquainted with i_4. We leave for future research a detailed exploration of how to combine temporal *de re*, indices and concept generators.

(1) $[\![\text{CG}_{ana,i_3}\ [\text{pro}_4]]\!]^g(i_3)$ = the index <w′,t′> such that: w′=w(i_3), t(i_3) < t′ and t′ has property α at i_3

[5]We leave the two occurrences of pro_6 free and co-indexed in (21), but of course they could also be not co-indexed or each existentially bound (via an $\exists$-closure operator). This will not play a role for our purposes.

(21) LF: [λ0 Ana think at $\text{pro}_1^{[\text{PAST pro}_0]}$ [λ2 $\text{MODAL}^{\text{L}}_{\text{METAPHY}}$ pro_2
[λ3 ~~past~~ [$\text{pro}_4^{[\text{FUT pro}_3]}$ [λ5 [$\text{pro}_6^{[\text{PAST pro}_5]}$ λ7[John go at pro_7]]]]]
[λ3 ~~past~~ [$\text{pro}_4^{[\text{FUT pro}_3]}$ [λ5 [$\text{pro}_6^{[\text{PAST pro}_5]}$ λ7[the party be a fun at pro_7]]]]]]]

(22) a. Antecedent clause: λi_3: $i_3 < i_4 \wedge i_6 < i_4$. John go at i_6
b. Consequent clause: λi_3: $i_3 < i_4 \wedge i_6 < i_4$. the.party be fun at i_6
c. Sentence: λi_0: $i_1 < i_0$. $\forall i_2 \in \text{Dox}_{\text{Ana}}(i_1)$ $\forall i_3 \in \text{Metaph}^{\text{L}}$ (i_2):
$i_3 < i_4 \wedge i_6 < i_4 \wedge$ John go at i_6 $\rightarrow$
$i_3 < i_4 \wedge i_6 < i_4 \wedge$ party be fun at i_6

3 Subjunctive mood

In Spanish and other Romance languages, representational verbs like *creer* 'believe' and *decir* 'say' select indicative mood in their complement clause, whereas non-representational verbs like *lamentar* 'regret' and *hacer* 'to make (somebody do something)' select subjunctive: (23)-(24).

(23) Bea cree [que Juan enseña / *enseñe semántica]
Bea believes [that Juan teaches.IND / *teaches.SUBJ semantics]
'Bea believes that Juan teaches semantics.'

(24) Bea lamenta [que Juan *enseña / enseñe semántica]
Bea regrets [that Juan *teaches.IND / teaches.SUBJ semantics]
'Bea regrets that Juan teaches semantics.'

[Sch05] analyses mood morphology as introducing mood features on world pronouns, as illustrated in (25). The features IND(icative) and SUBJ(unctive) are relative to an anchor attitude holder pro_k and call up the so-called "local context" pertaining to that attitude holder [Sta75], that is, the set of doxastic alternatives $\text{Dox}_{g(k)}$ of pro_k at the relevant evaluation world. The feature IND imposes the presupposition that the world variable of the verb is a member of that $\text{Dox}_{g(k)}$, whereas the feature SUBJ imposes no presupposition, as defined in (26)-(27):

(25) LF of the indicative morphology in a verbal form: $\text{pro}_i^{[\text{IND } pro_k]}$

(26) $[\![pro_i^{[\text{IND } pro_k]}]\!]$ is defined only if g(i) $\in \text{Dox}_{g(k)}$;
if defined, $[\![pro_i^{[\text{IND } pro_k]}]\!] = g(pro_i)$

(27) $[\![pro_i^{[\text{SUBJ } pro_k]}]\!] = g(pro_i)$

Let us see what happens when these mood features combine with the rest of the material in a clause. IND makes the resulting proposition partial, defined only for worlds w′ in $\text{Dox}_{g(k)}$, where g(k) is the referent x of the matrix subject. This is shown in (28). SUBJ yields the (usual) total proposition. To make clear that the presupposition is waived by having explicitly used a subjunctive form, we cross out the content of that presupposition in our formulas, as in (29):

(28) $[\![Juan\ teach\ semantics\ at\ pro^{[\text{IND } pro_k]}]\!] = \lambda w'{:}w' \in \text{Dox}_{g(k)}(w_0)$.J teaches sem in w′
= the function f such that, for any w in W:
f(w)=1 if w $\in \text{Dox}_{g(k)}(w_0)$ and John teaches semantics in w
f(w)=0 if w $\in \text{Dox}_{g(k)}(w_0)$ and John does not teach semantics in w and
f(w)=# if w $\notin \text{Dox}_{g(k)}(w_0)$

(29) $[\![\textit{Juan teach semantics at } pro_i^{[\text{SUBJ } pro_k]}]\!] = \lambda w'$:~~$w' \in Dox_{g(k)}(w_0)$~~.J teaches sem in w′

Now we are ready to combine the indicative and subjunctive complement clauses with the embedding verbs. We start with 'believe', defined in (30). This lexical entry simply asks us to check the value of our proposition at the worlds $w \in Dox_x(w_0)$. For that, the partial indicative proposition (28) suffices. By Maximize Presupposition in (31) [Hei91], indicative has to be used under 'believe' and, thus, a subjunctive complement clause is ungrammatical.

(30) $[\![\textit{believe}]\!](p)(x) = \lambda w_0.\ \forall w \cap Dox_x(w_0)\colon p(w)$

(31) Maximize Presupposition: Make your contribution presuppose as much as possible!

In the case of 'regret', defined in (32) [Hei92], it is presupposed that the subject x believes the proposition p, that is, that in all the worlds $w \in Dox_x(w_0)$, p is true at w. Then, for each such world w, $Sim_w(p)$ asks us to find, on the one hand, the most similar world w′ to w for which p(w′) yields TRUE/1 –which will be w itself– and $Sim_w(\neg p)$ ask us to find, on the other hand, the most similar world w′ to w for which ¬p(w′) yields TRUE/1. If we use the partial indicative proposition (28), the latter task cannot be carried out. Given that the subject x believes p, ¬p is the proposition in (33). But there is no world w′ –no matter how similar or dissimilar to w– for which ¬p(w′) yields TRUE/1. This means that the value of $Sim_w(\neg p)$ is undefined, which in turn means that sentence (24) with indicative leads to a presupposition failure. Since this presupposition failure arises systematically from the logical structure of the components, the sentence is ungrammatical (cf. [Gaj02]). If, instead, the total subjunctive proposition (29) is used, no presuppositional failure arises and the sentence is grammatical.

(32) $[\![\textit{regret}]\!](p)(x) = \lambda w_0\colon \forall w \cap Dox_x(w_0)\ [p(w)].$
$\forall w \cap Dox_x(w_0)\ [Sim_w(\neg p) >_{Bou_x(w0)} Sim_w(p)]$

(33) The function f such that, for any w in W:
$f(w)=0$ if $w \in Dox_x(w_0)$ and John teaches semantics in w
~~$f(w)=1$ if $w \in Dox_x(w_0)$ and John does not teach semantics in w~~ and
$f(w)=\#$ if $w \notin Dox_x(w_0)$

4 Proposal

[Dud83]'s original idea is that a counterfactual with 'fake' tense involves a back shift in time with a future (metaphysical) conditional interpreted under that back shift. Translating this idea into an LF structure gives us an interpretable past tense scoping over an entire future metaphysical conditional. Adding to that the analyses of tense and mood in the preceding sections, we obtain the following preliminary LFs for our present CC (34) (=(3)) and our past CC (36) (=(4)):

(34) Si Juan tuviese resaca (ahora/hoy), (pro) estaría en la cama. PRESENT
If Juan **had.**SUBJ hang-over (now/today), (he) **would-be** in the bed
'If John had a hang-over (now/today), he would be in bed.'

(35) Prelimary LF for present CC (34):
$\lambda 0\ [pro_1^{[\text{PAST } pro_0]}\ \lambda 2\ \text{MODAL}^{\text{L}}_{\text{METAPHY}}\ pro_2$
$[\lambda 8[pro_8^{[\text{SUBJ } pro_{Sp}]}\ \lambda 3\ [\ \sout{past}\ [pro_4^{[\text{FUT } pro_3]}\ \lambda 7[\text{John have hang-over at } pro_7]]]]]$
$[\lambda 8[pro_8\ \lambda 3\ [\ \sout{past}\ [pro_4^{[\text{FUT } pro_3]}\ \lambda 7[\text{John be in bed at } pro_7]]]]]\ ]$

(36) Si Juan hubiese ido a la fiesta (ayer), la fiesta habría sido
If Juan **had.**SUBJ **gone** to the party (yesterday), the party **would.have been**
divertida. PAST
amusing.

'If John had gone to the party (yesterday), the party would have been fun.'

(37) Prelimary LF for past CC (36):
$\lambda 0$ [$\text{pro}_1^{[\text{PAST pro}_0]}$ $\lambda 2$ $\text{MODAL}^{\text{L}}_{\text{METAPHY}}$ pro_2
[$\lambda 8$[$\text{pro}_8^{[\text{SUBJ pro}_{Sp}]}$ $\lambda 3$ [~~past~~ [$\text{pro}_4^{[\text{FUT pro}_3]}$ $\lambda 5$ [$\text{pro}_6^{[\text{PAST pro}_5]}$ $\lambda 7$[John go at pro_7]]]]]]
[$\lambda 8$[pro_8 $\lambda 3$ [~~past~~ [$\text{pro}_4^{[\text{FUT pro}_3]}$ $\lambda 5$ [$\text{pro}_6^{[\text{PAST pro}_5]}$ $\lambda 7$[it be fun at pro_7]]]]]]]

Two adjustments are still needed to derive appropriate truth conditions for CCs.

First, [Dud83]'s original idea needs to be profiled a bit more in order to guarantee the correct temporal alignment of the hypothetical events with respect to the utterance index i_0. To see this, recall the analysis of the indirect speech examples (10) and (13). There, the attitude holder Ana was thinking about how things would be on a particular date, represented in our LFs as a free pronoun pro_4 which happened to pick today's date in our scenarios. But of course Ana's thoughts could be about any other salient date given the appropriate scenario. When we turn to CCs, pronoun pro_4 cannot pick a random date but must be co-valued with pro_0. This is because, in the LF (35), the index $[\![pro_7]\!]/[\![pro_4]\!]$ at which John has a hang-over and John is in bed is necessarily understood as temporally overlapping with the utterance index $[\![pro_0]\!]$ and, in the LF (37), the index $[\![pro_7]\!]/[\![pro_6{}^{[\text{PAST } pro_5/pro_4]}]\!]$ at which John goes to the party and the party is fun is obligatorily interpreted as temporally preceding the utterance index $[\![pro_0]\!]$. This means that Dudman's original idea should be refined as follows: A (present or past) counterfactual uttered at index i_0 involves a back shift in time with a future metaphycial conditional *about* i_0 under that back shift. In other words, a counterfactual is not just about considering how the world would be like at some time if certain things had happened, but about how the world would be *now* is those things had happened. For concreteness, this is implemented by adding the feature [T-IDENT pro_0] to pro_4, interpreted as follows:

(38) $[\![pro_i^{[\text{T-IDENT } pro_j]}]\!]$ is defined only if time(g(i))=time(g(j));
if defined, $[\![pro_i^{[\text{T-IDENT } pro_j]}]\!]$ = g(i)

Second, it is known that CCs do not quantify over all future metaphysical possibilities branching out from a given past time t'. "Intermediate" facts that took place between t' and t_0 are sometimes taken into account too, with the effect that they further restrict the metaphysical possibilities quantified over. These are the so-called Morgenbesser cases, one example of which, from [Edg04], is given in (39):

(39) I am driving to the airport to catch a 9 o'clock flight to Paris. The car breaks down in the motorway. I sit there waiting for the breakdown service. 9 o'clock passes: I've missed my flight. More time passes. 'If I had caught the plane, I would have been half way to Paris by now', I say to the repairman who eventually shows up. 'Which flight were you on?', he asks. I tell him. 'Well you're wrong', he says. 'I was listening to the radio. It crashed. If you had caught that plane, you would be dead by now.'

Unless we want to commit to determinism (and assume that the later plane crash at t'' was already determined at t'), this case makes clear that the domain of quantification of the modal cluster [$\text{MODAL}^{L}_{\text{METAPHY}}$ pro_2] in our LFs should not include all the metaphysical possibilities

branching out from index $[\![pro_2]\!]$ that follows the set of laws L, but only the possibilities out of those that, additionally, contain certain later facts –here, the plane crash. While there is important discussion in the literature on how to characterize which intermediate facts play a role and which ones do not, our modest goal here is to have a place holder for that information in our LFs. For concreteness, we implement this by adding a situation argument pro_{sit} to the modal cluster, whose denotation should be a modal part ($\subseteq_m$) of w′, as defined in (40) [Arr09]. Our complex modal cluster [$\text{MODAL}^{L}_{\text{METAPHY}}$ pro_{sit} pro_2] is interpreted as in (41):

(40) For any situation s and world w:
s $\subseteq_m$ w iff there is a situation s′ such that s′ is a counterpart of s and s′ is part of w.

(41) $[\![\text{MODAL}^{L}_{\text{METAPHY}}\ pro_{sit}\ pro_2]\!](\text{p})(\text{q}) =$
λi. $\forall$i′ $\in$ (Metaph^{L}(i) $\cap$ {<w′,t′>:$[\![pro_{sit}]\!] \subseteq_m$ w′}) [p(i′) $\rightarrow$ q(i′)]

Let us add these two adjustments to our preliminary LFs. For the present CC (34), we obtain the LF (42). This leads to the partial semantic computation in (43). The resulting truth conditions (43c) quantify over law-like metaphysical alternatives i_3 to an index i_2 preceding the utterance index i_0, alternatives at which, additionally, certain "intermediate" facts hold. For each of these i_3, we check whether the index i_4 that has the same world-member as i_3 and the same time-member as i_0 is such that John has a hang-over at i_4.[6] If so, then the sentence commits us to i_4 being such that John is in bed at i_4. This delivers the correct temporal alignment of the hypothetical events. As for mood, the use of subjunctive in the *if*-clause makes the antecedent proposition total, as in (43a). If, instead, indicative mood were used, the antecedent proposition would be defined only for the doxastic alternatives of the attitude holder, here the speaker. Since the speaker believes that the antecedent is false, this would lead to vacuous quantification: For any index i_3 that we would apply the indicative version of (43a) to, we would obtain # (if $i_3 \notin \text{Dox}_{Speaker}(i_0)$) or FALSE/0 (if $i_3 \in \text{Dox}_{Speaker}(i_0)$). Hence, indicative mood cannot be used and subjunctive mood must.

(42) LF for present CC (34):
λ0 [$\text{pro}_1^{[\text{PAST pro}_0]}$ λ2 $\text{MODAL}^{\text{L}}_{\text{METAPHY}}$ (pro_{sit}) pro_2
[λ8[$\text{pro}_8^{[\text{SUBJ pro}_{Sp}]}$ λ3[~~past~~ [$\text{pro}_4^{[\text{FUT pro}_3][\text{TEMP pro}_0]}$ λ7[John have hang-over at pro_7]]]]]
[λ8[pro_8 λ3[~~past~~ [$\text{pro}_4^{[\text{FUT pro}_3][\text{TEMP pro}_0]}$ λ7[John be in bed at pro_7]]]]]]

(43) a. Antecedent clause:
λi_3: ~~$i_3 \in \text{Dox}_{Sp}(i_0)$~~ $\wedge$ $i_3 < i_4$ $\wedge$ time(i_4)=time(i_0). John have hang-over at i_4

b. Consequent clause:
λi_3: $i_3 < i_4$ $\wedge$ time(i_4)=time(i_0). John be in bed at i_4

c. Sentence:
λi_0: $i_1 < i_0$. $\forall i_3 \in$ ($\text{Metaph}^{L}(i_1) \cap$ {<w′,t′>:$[\![pro_{sit}]\!] \subseteq_m$ w′}):
~~$i_3 \in \text{Dox}_{Sp}(i_0)$~~ $\wedge$ $i_3 < i_4$ $\wedge$ time(i_4)=time(i_0) $\wedge$ John have hang-over at i_4 $\rightarrow$
$i_3 < i_4$ $\wedge$ time(i_4)=time(i_0) $\wedge$ John be in bed at i_4

For the past CC (36), we obtain the following LF and semantic derivation. As the reader can check for herself, now the index i_6 at which the hypothetical events of the antecedent and consequent clauses hold has to temporally precede i_4. Again, i_4 has the same time-member as i_0 (and the same world-member as i_3). This leads to the correct temporal ordering. As for mood, the same considerations apply as above.

[6] See footnote 4 on i_4. Alternatively, one could $\exists$-bind i_4.

(44) LF for past CC (36):
$\lambda 0$ [$\text{pro}_1^{[\text{PAST pro}_0]}$ $\lambda 2$ $\text{MODAL}^{\text{L}}_{\text{METAPHY}}$ (pro_{sit}) pro_2
[$\lambda 8$[$\text{pro}_8^{[\text{SUBJ pro}_{Sp}]}$ $\lambda 3$[~~past~~ [$\text{pro}_4^{[\text{FUT pro}_3][\text{TEMP pro}_0]}$ $\lambda 5$[$\text{pro}_6^{[\text{PAST pro}_5]}$ $\lambda 7$[J go at pro_7]]]]]]
[$\lambda 8$[pro_8 $\lambda 3$[~~past~~ [$\text{pro}_4^{[\text{FUT pro}_3][\text{TEMP pro}_0]}$ $\lambda 5$[$\text{pro}_6^{[\text{PAST pro}_5]}$ $\lambda 7$[it be fun at pro_7]]]]]]

(45) a. Antecedent clause:
λi_3: ~~$i_3 \in \text{Dox}_{Sp}(i_0)$~~ $\wedge$ $i_3 < i_4$ $\wedge$ time(i_4)=time(i_0) $\wedge$ $i_6 < i_4$. John go at i_6

b. Consequent clause:
λi_3: $i_3 < i_4$ $\wedge$ time(i_4)=time(i_0) $\wedge$ $i_6 < i_4$. the.party be fun at i_6

c. Sentence:
λi_0: $i_1 < i_0$. $\forall i_3 \in (\text{Metaph}^L(i_1) \cap \{<w',t'>:[\![pro_{sit}]\!] \subseteq_m w'\})$:
~~$i_3 \in \text{Dox}_{Sp}(i_0)$~~ $\wedge$ $i_3 < i_4$ $\wedge$ time(i_4)=time(i_0) $\wedge$ $i_6 < i_4$ $\wedge$ John go at i_6 $\rightarrow$
$i_3 < i_4$ $\wedge$ time(i_4)=time(i_0) $\wedge$ $i_6 < i_4$ $\wedge$ the.party be fun at i_6

In sum, the correct truth conditions have been derived for our CC using the analysis of tense and mood morphology independently motivated in sections 2 and 3.

5 Conclusions and further issues

The truth conditions of CCs in Spanish have been derived within the temporal remoteness line while keeping a uniform analysis of temporal and mood morphology across the grammar.

I would like to make two further points about the temporal remoteness approach.

First, relating CC structures to the description of future events under a back shift cannot only account for 'fake' tense, as we saw, but also for 'fake' aspect. It has been noted that, even when the event described in the antecedent clause is punctual, an imperfective past form has to be used. We note that the same is true for indirect speech reporting a past utterance of an indicative future conditional. To see one case, our indirect speech example (10) uses Past Imperfect (= Pretérito Imperfecto) in the antecedent clause. If we use Perfective Past (= Pretérito Indefinido) in (46), the sentence lacks a parallel interpretation.

(46) Ella pensó que, si Juan alcanzó la cima, (pro) estaría cansado.
She thought that, if Juan **reached.Pft.Ind** the summit, he **would.be** tired

Second, counterpossibles like *[If 2 plus 2 were 5, . . .]* have always been an important problem for the temporal remoteness line. While I have no real solution to offer, one possible avenue to explore is to relativize indicative and counterfactual conditionals to a given epistemic state (cf. [Sta14, Lea17]). In that case, Dudman's back shift may be understood not as taking us back to a time point t′ at which the metaphysical future conditional is true, but to a time point t′ at which the some agent's epistemic state deems the metaphysical future conditional true.

References

[Abu97] Dorit Abusch. Sequence of tense and temporal de re. *Linguistics and Philosophy*, 20:1–50, 1997.

[And51] Alan Ross Anderson. A note on subjunctive and counterfactual conditionals. *Analysis*, 11:35–38, 1951.

[Arr09] Ana Arregui. On similarity in conditionals. *Linguistics and Philosophy*, 32:245–278, 2009.

[Dud83] V. H. Dudman. Tense and time in english verb clusters of the primary pattern. *Australian Journal of Linguistics*, 3:25–44, 1983.

[Dud84] V. H. Dudman. Conditional interpretations of if-sentences. *Australian Journal of Linguistics*, 4:143–204, 1984.

[Edg04] Dorothy Edgington. Counterfactuals and the benefit of hindsight. In Phil Dowe and Paul Noordhof, editors, *Cause and chance: Causation in an indeterministic world*, pages 143–170. Rouledge, London, 2004.

[Gaj02] Jon Gajewski. L-analycity in natural language. Ms. U. Connecticut, 2002.

[GvS09] Atle Grønn and Arnim von Stechow. Temporal interpretation and organization of subjunctive conditionals. Ms. U. Oslo, 2009.

[Hei91] Irene Heim. Artikel und definitheit. In A. von Stechow and D. Wunderlich, editors, *Semantik: Ein internationales Handbuch der zeitgenoessischen Forschung*, pages 487–535. de Gruyter, Berlin, 1991.

[Hei92] Irene Heim. Presupposition projection and the semantics of attitude verbs. *Journal of Semantics*, 9:183–221, 1992.

[Hei94] Irene Heim. Comments on abusch's theory of tense. In H. Kamp, editor, *Ellipsis, Tense and Questions*, pages 143–170. U. Amsterdam, 1994.

[Iat00] Sabine Iatridou. The grammatical ingredients of counterfactuality. *Linguistic Inquiry*, 31:231–270, 2000.

[Ipp03] Michela Ippolito. Presuppositions and implicatures in counterfactuals. *Natural Language Semantics*, 11:145–186, 2003.

[Ipp13] Michela Ippolito. *Subjunctive conditionals: A linguistic analysis*. MIT Press, Cambridge, MA, 2013.

[Kau05] Stefan Kaufmann. Conditional truth and future reference. *Journal of Semantics*, 22:231–280, 2005.

[Kra98] Angelika Kratzer. More structural analogies between pronouns and tenses. In D. Strolovitch and A. Lawson, editors, *Proceedings of SALT 8*, Ithaca, N.Y., 1998. CLC Publications.

[Kus05] Kiyomi Kusumoto. On the quantification over times in natural language. *Natural Language Semantics*, 13:317–357, 2005.

[Lea17] Brian Leahy. Counterfactual antecedent falsity and the epistemic sensitivity of counterfactuals. *Philosophical Studies*, pages 1–25, 2017.

[Lew73] David Lewis. *Counterfactuals*. Blackwell, Malden, 1973.

[Par73] Barbara Partee. Some structural analogies between tenses and pronouns in english. *Journal of Philosophy*, 7:601–609, 1973.

[PS03] Orin Percus and Uli Sauerland. On the lfs of attitude reports. In M. Meisgerber, editor, *Proceedings of Sinn und Bedeutung 7*, Konstanz, 2003. University of Konstanz.

[Rom14] Maribel Romero. 'fake tense' in counterfactuals: A temporal remoteness approach. In Luca Crnic and Uli Sauerland, editors, *The Art and Craft of Semantics: a Festschrift for Irene Heim*, volume 2, pages 47–63. MITWPL, Cambridge, MA, 2014.

[Sch05] P. Schlenker. The lazy frenchman's approach to the subjunctive. In *Proceedings of going Romance XVII*, 2005.

[Sch14] Katrin Schulz. Fake tense in conditional sentences: A modal approach. *Natural Language Semantics*, 22:117–144, 2014.

[Sta75] R. Stalnaker. Indicative conditionals. *Philosophia*, 5:269–286, 1975.

[Sta14] Robert Stalnaker. *Context*. Oxford University Press, Oxford, 2014.

[vS95] A. von Stechow. On the proper treatment of tense. In M. Simons and T. Galloway, editors, *Proceedings of SALT 5*, Ithaca N.Y., 1995. CLC Publications.

[vS09] A. von Stechow. Tense in compositional semantics. To appear in W. Klein, ed., *The Expression of Time*. De Gruyter. Available at: http://www.sfs.uni-tuebingen.de/ astechow/Aufsaetze/Approaches.pdf, 2009.

Conditional Excluded Middle in Informational Semantics*

Paolo Santorio

University of California, San Diego, San Diego CA, USA
psantorio@ucsd.edu

Abstract

Semantics for indicative conditionals (ICs) struggle with a problem inherited from the classical Stalnaker/Lewis debate on counterfactuals. On the one hand, ICs seem to satisfy Conditional Excluded Middle; on the other, ICs of the form $\phi > \neg\psi$ seem incompatible with *might*-conditionals of the form $\phi > \Diamond\psi$. These requirements are jointly unsatisfiable on standard notions of consequence. I show that a relative of Veltman's data and update semantics (1985, 1996), which I call *path semantics*, validates both. The analysis is confined to ICs, but can in principle be extended to counterfactuals.

1 Introduction

All theories of conditionals struggle with a tension between two plausible logical principles, which is inherited from the classical debate on counterfactuals between Stalnaker ([19]) and Lewis ([15]). On the one hand, conditionals seem to satisfy Conditional Excluded Middle, i.e. the principle that sentences of the form $(\phi > \psi) \vee (\phi > \neg\psi)$ are valid. On the other, conditionals of the form $\phi > \neg\psi$ seem incompatible with *might*-conditionals of the form $\phi > \Diamond\psi$. Unfortunately, these requirements are jointly unsatisfiable on a classical notion of consequence. As a result, most theories of conditionals drop one of them.

I show that the tension can be solved by moving to a new semantics that generates a nonclassical notion of consequence, which I call *path semantics*. Path semantics is a relative of informational semantics for epistemic modality in the style of Veltman ([20], [21]). In path semantics, all sentences are evaluated as true and false at sequences of information states. Conditionals have no quantificational force; rather, their antecedents are used to update the sequence of evaluation. Path semantics generates nonclassical notions of consequence that vindicate both the logical principles at stake.

I proceed as follows. §2 sets up the background problem; §3 briefly discusses solutions based on homogeneity; §4 introduces path semantics, and §5 discusses consequence. Given space constraints, throughout the paper I focus on epistemic conditionals and their *might*-counterparts, though both the puzzle and the account can be generalized to counterfactuals.

2 The puzzle

2.1 Conditional Excluded Middle

The first principle I consider is Conditional Excluded Middle (below), which I defend via two lines of argument.

Conditional Excluded Middle. (CEM) $\models (\phi > \psi) \vee (\phi > \neg\psi)$

*Thanks to Paul Egré, Manuel Križ, Matt Mandelkern, Salvador Mascarenhas, Benjamin Spector, to three Amsterdam Colloquium referees, and to audiences at the 2017 Dubrovnik workshop in philosophy of language and Institut Jean Nicod in Paris. Special thanks to Maria Aloni and Frank Veltman, whose questions during a talk on related material led to the formulation of these ideas. All mistakes remain mine.

Argument #1: scopelessness. Epistemic conditionals with no overt modal appear to be scopeless with respect to logical operators: importing and exporting these operators inside and outside the consequent of a conditional makes no difference to truth conditions. For reasons of space here I only discuss negation, but the evidence for scopelessness includes the interactions between conditionals and quantifiers ([8], [5]), the adverb *only* ([4]), and comparative constructions ([10]).

Notice, first of all, that negation can be imported inside and outside the scope of a conditional without affecting truth conditions. The sentences in (1) are equivalent.

(1) a. It's not the case that, if Frida took the exam, she passed.
b. If Frida took the exam, she didn't pass.

Notice also that the phenomenon persists with items that lexicalize negation, like *doubt* (≈ *believe not*) and *fail* (≈ *not pass*).

(2) a. I doubt that, if Frida took the exam, she passed.
b. I believe that, if Frida took the exam, she failed.

The lack of semantic interaction with negative items is perfectly expected on a theory that vindicates CEM (assuming standard Excluded Middle in the background logic), but not on theories that treat conditionals as universal quantifiers.

Argument #2: probability. *Modulo* plausible assumptions, CEM is needed to vindicate basic probability judgments about conditionals. Both intuition and experimental results suggest that speakers judge that, at least for unembedded conditionals, the probability of a conditional equals the conditional probability of the consequent, given the antecedent.[1]

The Thesis. $Pr(\phi > \psi) = Pr(\psi \mid \phi)$
(for all ϕ, ψ, and for all Pr modeling rational credence)

Unfortunately, as has been shown repeatedly in the literature on so-called triviality results (see a.o. [14], [7]), the Thesis is untenable in full generality in truth-conditional frameworks. At the same time, one reasonable goal for semantic theory is to give a partial vindication of the Thesis. A plausible semantic theory should allow for assigning probabilities that conform to the Thesis to most simple conditionals in most ordinary contexts (for this claim, see e.g. [17]).

This piecemeal vindication seems highly desirable. Now, given some plausible assumptions, even this modest goal forces the adoption of CEM. Assume that, if a conditional $\phi > \psi$ conforms to the Thesis, then the corresponding conditional with a negated consequent $\phi > \neg\psi$ also conforms to the Thesis.[2] We can prove that, whenever $\phi > \psi$ has a consistent antecedent and conforms to the Thesis, the corresponding instance of CEM has probability 1.[3] In short:

[1] For a survey of classical experimental literature, see [3].

[2] Whatever one thinks about the Thesis in general, this assumption seems particularly intuitive, at least for simple conditionals. Even alleged counterexamples to the Thesis (see e.g. [9]) seem to conform to this assumption.

[3] Assuming the following uncontroversial principle, the proof is below.

Conditional noncontradiction. (CNC) $(\phi > \psi) \supset \neg(\phi > \neg\psi)$

i. $Pr(\psi \mid \phi) + Pr(\neg\psi \mid \phi) = 1$ (probability calculus)
ii. $Pr(\phi > \psi) + Pr(\phi > \neg\psi) = 1$ (i, Thesis)
iii. $Pr((\phi > \psi) \wedge (\phi > \neg\psi)) = 0$ (CNC)
iv. $Pr((\phi > \psi) \vee (\phi > \neg\psi)) = 1$ (ii, iii, probability calculus)

Fact. For all clauses ϕ, ψ (with ϕ consistent), and probability function Pr, such that $Pr(\phi > \psi) = Pr(\psi \mid \phi)$ and $Pr(\phi > \neg\psi) = Pr(\neg\psi \mid \phi)$, $Pr((\phi > \psi) \vee (\phi > \neg\psi)) = 1$.

Strictly speaking, Fact doesn't requre CEM; all we need is that a (large) number of instances of CEM are assigned probability 1. But a semantics that validates CEM immediately satisfies Fact. Conversely, accommodating Fact is a substantial challenge for any semantics that doesn't validate CEM.[4]

2.2 *If* and *might*

The second principle at stake states the incompatibility of $\phi > \neg\psi$ and $\phi > \Diamond\psi$.

If-Might Contradiction. (IMC) $(\phi > \neg\psi) \wedge (\phi > \Diamond\psi) \vDash \bot$

The evidence for IMC is straightforward. Discourses that involve conditionals of both forms are standardly heard as inconsistent; moreover, pairs of conditionals of this form can be used to generate disagreement. In addition, this infelicity persists also in linguistic environments that screen off pragmatic clashes, like supposition contexts (see [22]).

(3) # If Maria passed, Frida didn't pass; but, even if Maria passed, it might be that Frida passed.

(4) A: If Maria passed, Frida didn't pass.
B: I disagree. Even if Maria passed, it might be that Frida passed.

(5) # Suppose that, if Maria passed, Frida didn't pass, and that, if Maria passed, it might be that Frida passed.

Notice that IMC should be kept distinct from the following:

Duality. $\vDash (\phi > \Diamond\psi) \leftrightarrow (\neg(\phi > \neg\psi))$

Several classical frameworks (e.g., [15] [11], [12]) make IMC and Duality equivalent. But, as I show in §5, the two can come apart.

2.3 Collapse

Given a classical notion of consequence, CEM and IMC together entail the equivalence of $\phi > \psi$ and $\phi > \Diamond\psi$. The direction $\phi > \psi \vDash \phi > \Diamond\psi$ is uncontroversial; as for the other direction:

i.	$\phi > \Diamond\psi$	Assumption
ii.	$\quad\phi > \neg\psi$	Supposition for conditional proof
iii.	$\quad\phi > \neg\psi \wedge \phi > \Diamond\psi$	(i, ii, $\wedge$-Introduction)
iv.	$\quad\bot$	(iii, IMC)
v.	$\neg(\phi > \neg\psi)$	(ii-iv, Reductio)
vi.	$\phi > \neg\psi$	(v, CEM, Disjunctive syllogism)

Of course, this result is unacceptable. In response, classical theories drop one of CEM and IMC. Famously, Stalnaker ([19]) endorses CEM and rejects IMC, while most other theorists, ranging from Kratzer ([11] [12]) to Gillies ([6]) reject CEM. Both solutions are empirically costly, as is suggested by the discussion in this section.

[4] For example, on a Lewis/Kratzer-style semantics we need to make sure that, in each context, the ordering we use is appropriately tailored; this is bound to generate particularly implausible results in plenty of cases.

3 Homogeneity?

Some theorists ([4], [18], a.o.) try to capture both ICM and CEM by assuming that conditionals give rise to a so-called homogeneity inference. For concreteness I consider Schlenker's account, though my discussion generalizes to any homogeneity-based semantics. Before proceeding, a *caveat*. Both von Fintel and Schlenker treat homogeneity as a presupposition. Recently, Križ ([13]) has convincingly argued that homogeneity effects in natural language are not presuppositional. Here I follow Križ, though nothing I say depends on this.

Schlenker's account is based on an extended analogy between conditionals and plural definite descriptions. Roughly, conditionals are analyzed as modal descriptions of antecedent worlds. Here is an informal gloss of Schlenker's truth conditions:[5]

> $\ulcorner \phi > \psi \urcorner$ is true at w iff the closest ϕ-worlds to w are ψ-worlds

Schlenker notices that this semantics vindicates a version of CEM when supplemented with homogeneity. It is well-known that plural descriptions like *the books* give rise to a homogeneity inference. Roughly, this is the requirement that sentences involving plural descriptions have a determinate truth value just in case all the atoms in the denotation of the description behave in the same way with respect to the predicate. For example, (6) is determinately true or determinately false just in case Mary either has read all the books or has read none.

(6) Mary read the books.

For current purposes, it's not important how homogeneity is triggered. What matters is that, if we pursue the analogy with descriptions, we expect conditionals to trigger a similar inference.

> **Homogeneity Inference (HI)**
> $\phi > \psi$ is true or false at w only if: either all closest ϕ-worlds to w are ψ-worlds, or all closest ϕ-worlds to w are $\neg\psi$-worlds

Given background assumptions about negation, HI allows us to vindicate a version of CEM.[6] *Modulo* a plausible semantics for *might*, IMC is also vindicated while avoiding collapse.[7]

While this is at first sight promising, there are reasons to reject the close analogy between plural descriptions and conditionals. Here I mention two.

Disanalogy #1: Probability. HI is of no help in vindicating intuitions about probability (see [1] for a similar argument about *will*). Consider:

(7) If Maria flipped the coin, the coin came up tails.

Suppose that you have no evidence one way or the other about how the coin landed. Plausibly, then, flip-and-heads-worlds and flip-and-tails-worlds are tied for closeness in your epistemic state.[8] In this situation, it seems that you should assign probability .5 to (7). Yet, if we assume

[5] On this gloss, as well as on Schlenker's theory, both conditionals and plural descriptions are taken to be nonmonotonic. As Schlenker makes clear, this is a non-essential feature of the theory and may be dropped.

[6] The background assumption is that negation has a Strong Kleene semantics, taking falsity into truth. Notice also that Schlenker's semantics, supplemented with HI, vindicates a weaker principle than CEM proper: $(\phi > \psi) \vee (\phi > \neg\psi)$ is valid *whenever* $\phi > \psi$ *is not undefined*

[7] Treat $\phi > \Diamond\psi$ as involving existential quantification over the closest ϕ-worlds. Then $\phi > \Diamond\psi$ is incompatible with $\phi > \neg\psi$, but the entailment to $\phi > \psi$ is blocked. To be sure, we have that, whenever $\phi > \Diamond\psi$ is true, $\phi > \neg\psi$ is not false; but it can be that $\phi > \Diamond\psi$ is true and $\phi > \neg\psi$ is undefined.

[8] I assume that the specific criteria we adopt for epistemic closeness won't matter here. If you think they do, just switch to a different example.

HI, (7) suffers from homogeneity failure and hence it is undefined. Now, it is unclear exactly what credence one should assign to statements with this status. But it seems both irrational and unusual to assign to them positive intermediate credence. For a comparison, ask yourself what credence you would assign to "Mary read the books" in a scenario where Mary read only half of the books.

This theoretical argument seems supported by recent experimental findings. Cremers, Križ and Chemla find ([2]) that subjects tend to follow different patterns when asked to assign probabilities, on the one hand, to sentences suffering from homogeneity failure and, on the other, to conditionals in contexts where not all possible antecedent worlds verify the consequent.

Disanalogy #2: Projection behavior. Homogeneity inferences project under attitude verbs. In particular, they display a projection behavior similar to that of presuppositions under attitude verbs or complex predicates that describe uncertainty, like *wonder*, or *be not certain*. For example, *S wonders whether p*, where *p* involves a homogeneity trigger, suggests the inference that *S* believes the relevant homogeneity claim.[9]

(8) a. Paula wonders whether the girls passed.
$\rightsquigarrow$ Paula believes: all girls passed $\vee$ all girls didn't pass
b. Everyone is not certain that the girls passed.
$\rightsquigarrow$ Everyone believes: all girls passed $\vee$ all girls didn't pass

Conversely, conditionals don't display a similar projection behavior.[10]

(9) a. Paula wonders whether, if Frida took the exam, she passed.
$\not\rightsquigarrow$ Paula believes: all her epistemic exam-worlds are pass-worlds $\vee$ all her epistemic exam-worlds are not-pass-worlds.
b. Everyone is not certain that Frida passed, if she took the exam.
$\not\rightsquigarrow$ Everyone believes: all their epistemic exam-worlds are pass-worlds $\vee$ all their epistemic exam-worlds are non-pass-worlds

(I'm assuming that conditionals in the scope of an attitude verb quantify over the subject's epistemic state. The problem could be reproduced on other accounts, e.g. accountswhere conditionals directly take their domain of quantification from attitude verbs, in the style of [22].)

4 Path semantics

The puzzle I outlined in §2 is due to the tension generated by three individually plausible but seemingly inconsistent principles.

Conditional Excluded Middle. (CEM)	$\vDash (\phi > \psi) \vee (\phi > \neg\psi)$
If-Might Contradiction. (IMC)	$(\phi > \neg\psi) \wedge (\phi > \Diamond\psi) \vDash \bot$
Nonfactivity of Might-Conditionals. (NMC)	$\phi > \Diamond\psi \not\vDash \phi > \psi$

When framed in this way, the puzzle is clearly reminiscent of a classical puzzle about epistemic *might*. Yalcin ([22]) notices that sentences of the form $\neg\phi \wedge \Diamond\phi$ seem inconsistent:

(10) # It's not raining and it might be raining.

[9] For the claim that *S wonders whether* Q_p presupposes *S believes p*, see [16].

[10] I assume that conditionals in attitude reports quantify over epistemic worlds of the subject of the attitude. The point still goes through if we assume, loosely following Yalcin ([22]), that conditionals merely borrow the domain of quantification from the relevant attitude verbs.

Yet this intuition is hard to vindicate on a classical semantics for *might*, on which the following two principles are inconsistent.

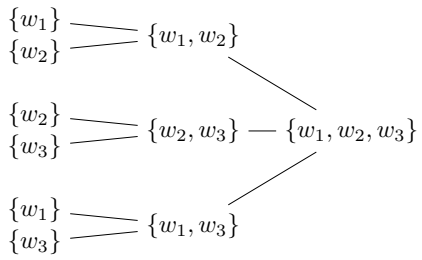
Epistemic Contradiction. $\neg\phi \wedge \Diamond\phi \vDash \bot$
Nonfactivity of Epistemic Modality. $\Diamond\phi \nvDash \phi$

The puzzle in §2 is then a generalization of Yalcin's puzzle (to see this, just instantiate $\top$ for 'ϕ' in our triad). Yalcin takes his puzzle as motivation for pursuing an informational semantics for epistemic modals, in the style of Veltman ([20], [21]). I pursue a similar goal. The new semantics will allow us to define a nonclassical notion of consequence, on which the three offending principles are consistent.

4.1 Paths

On standard informational semantics (see e.g. [20], [21]), sentences are evaluated as true or false relative to an *information state*, which here I model as a set of worlds. Conversely, the basic unit of path semantics is an *information path*. Informally, an information path is a sequence of information states that starts from the empty set and expands into a larger information state, adding one world at a time. More formally, we can define paths as follows. Let i be any information state. An information path (in i) is a sequence of subsets of i that is (i) ordered by subsethood and (ii) maximal, in the sense that there is no larger sequence of subsets of i that is ordered by subsethood. An information state uniquely determines a set of information paths; I use $\textsc{path}(i)$ to denote the set of paths determined by information state i.

It is useful to model paths via branching diagrams. For an example, here is the set of paths determined by the information state $i = \{w_1, w_2, w_3\}$. (To avoid clutter, I leave out the empty set, which is the beginning point of each path.)



Intuitively, paths model possible ways in which information may grow. This becomes clearer by reading paths from right to left. A move to a smaller set in a path represents a possible transition from a less informed to a more informed state. (I will keep writing paths from the smallest to the largest set, left to right, because it makes the formalism more intuitive.)

4.2 Truth and falsity at a path

I state a semantics for a propositional language involving atomic sentences, Boolean connectives, epistemic modals, and conditionals. All sentences are evaluated relative to a path. This evaluation procedure is supplemented with a notion of update, which plays a key role in evaluating conditionals. In the next paragraphs, I state the semantics and go through some examples.

4.3 Semantics

I take an **information path in** i to be a maximal sequence of elements of $\wp(i)$, ordered by the subset relation. I use the customary square brackets '$[\![\cdot]\!]$' notation for the interpretation function and relativize interpretation to a path parameter P. I also assume a background model $\langle W, V\rangle$, with W a set of worlds and V a valuation function mapping atomic sentences to $\{0, 1\}$.

These are the clauses for atomic sentences, connectives, and modals.

Atoms: $[\![A]\!]^P = 1$ iff $w : \mathsf{min}(P) = \{w\}$, is s.t. $V(w, A) = 1$
($\mathsf{min}(P)$ is the smallest non-empty member of P)

$[\![\neg\phi]\!]^P = 1$ iff $[\![\phi]\!]^P = 0$

$[\![\phi \vee \psi]\!]^P = 1$ iff $[\![\phi]\!]^P = 1$ or $[\![\psi]\!]^P = 1$

$[\![\phi \wedge \psi]\!]^P = 1$ iff $[\![\phi]\!]^P = 1$ and $[\![\psi]\!]^P = 1$

$[\![\Diamond\phi]\!]^P = 1$ iff for some $w \in \bigcup P$, $V(w, \phi) = 1$

$[\![\Box\phi]\!]^P = 1$ iff for all $w \in \bigcup P$, $V(w, \phi) = 1$

To give a semantics for conditionals, we first need to define the update $P[\phi]$ of a path P with a formula ϕ. To do this, we define two preliminary notions. The first:

ϕ is **true throughout** an information state i iff, for all P in $\textsc{path}(i)$, $[\![\phi]\!]^P = 1$

I.e.: ϕ is true throughout an information state i just in case it is true at all the paths that are generated by i. Second, we define the notion of the update of an information state with ϕ.

i' **is the update of** i **with respect to** ϕ (in short: $i[\phi]$) iff:

i. $i' \subseteq i$;

ii. ϕ is true throughout i';

iii. there is no larger set that meets conditions (i) and (ii).

In short: the update of i with respect to ϕ is the largest subset of i such that ϕ is true at all the paths generated by it.[11] It is easy to check that this yields intuitive results.[12]

Finally, we define the update of P with respect to ϕ. This is just pointwise intersection of each information state in P with the updated information state that generates P:

Update of ϕ **with respect to** ϕ: $P[\phi] = P \Cap (\bigcup P)[\phi]$

(with: $P \Cap i = \langle p_1 \cap i, \ldots, p_n \cap i \ldots\rangle$)

At this point, we can define truth at a path for conditionals in terms of update.

$[\![\text{if } \phi, \psi]\!]^P = 1$ iff $[\![\psi]\!]^{P[\phi]} = 1$

[11] Given the sentences we're able to express in the language, there will always be a unique such set.

[12] Some examples: for any nonmodal sentence ϕ, $i[\phi]$ is the set of ϕ-worlds in i; $i[\Diamond\phi]$ is i itself if i contains a ϕ-world, and $\varnothing$ otherwise; $i[\neg\phi \wedge \Diamond\phi]$ is invariably $\varnothing$. These predictions are in line with update semantics ([21]). An interesting, and in my view welcome, divergence: $i[\Box\phi]$ is identical to $i[\phi]$, i.e. i updated with $\Box\phi$ is the set of ϕ-worlds in i.

4.4 Examples

It is useful to see how a few sentences are evaluated at a sample path. For illustration, consider: ('M' and 'F' stand for the propositions that Maria passed and that Frida passed):

(11) $\langle \varnothing, \{w_{\bar{\text{M}}\bar{\text{F}}}\}, \{w_{\bar{\text{M}}\bar{\text{F}}}, w_{\text{M}\bar{\text{F}}}\}, \{w_{\bar{\text{M}}\bar{\text{F}}}, w_{\text{M}\bar{\text{F}}}, w_{\bar{\text{M}}\text{F}}\}, \{w_{\bar{\text{M}}\bar{\text{F}}}, w_{\text{M}\bar{\text{F}}}, w_{\bar{\text{M}}\text{F}}, w_{\text{MF}}\}\rangle$

Nonmodal sentences. Nonmodal sentences are invariably evaluated at the first non-empty set in a path, which by design contains a unique world. As a result, the semantics of nonmodal sentences is fully classical. Here are some examples of sentences that are true at (11).

(12)
a. Maria didn't pass.
b. Maria passed or Frida didn't pass.
c. Neither Maria nor Frida passed.

Modalized sentences. $\Diamond\phi$ and $\Box\phi$ are evaluated at the information state that generates the path: technically, this means that they are evaluated at the union set of the path. As a result, the semantics of modalized claims reduces to standard informational semantics. For illustration, (11) makes true:

(13)
a. It might be that Maria passed.
b. It might be that Frida didn't pass.
c. It might be that Maria didn't pass and Frida did.

Conditionals. Conditionals make full use of the path structure. Conditional antecedents update the path of evaluation; the consequent is evaluated at the updated path. Consider:

(14) If Maria passed, Frida passed.

We first use the antecedent *If Maria passed* to update the path. Starting from (11) and following the definition of update above, we get:

$$\begin{array}{ccccccccc} \langle \varnothing, & \{w_{\bar{\text{M}}\bar{\text{F}}}\}, & \{w_{\bar{\text{M}}\bar{\text{F}}}, w_{\text{M}\bar{\text{F}}}\}, & \{w_{\bar{\text{M}}\bar{\text{F}}}, w_{\text{M}\bar{\text{F}}}, w_{\bar{\text{M}}\text{F}}\}, & \{w_{\bar{\text{M}}\bar{\text{F}}}, w_{\text{M}\bar{\text{F}}}, w_{\bar{\text{M}}\text{F}}, w_{\text{MF}}\}\rangle \\ \downarrow & \downarrow & \downarrow & \downarrow & & \downarrow & \\ \varnothing & \varnothing & \{w_{\text{M}\bar{\text{F}}}\} & \{w_{\text{M}\bar{\text{F}}}\} & & \{w_{\text{M}\bar{\text{F}}}, w_{\text{MF}}\} & \end{array}$$

From here, removing redundancy:

(15) $\langle \varnothing, \{w_{\text{M}\bar{\text{F}}}\}, \{w_{\text{M}\bar{\text{F}}}, w_{\text{MF}}\}\rangle$

At this point, we evaluate the consequent at the updated path. *Frida passed* is false at (15), hence the conditional is false at (11). Notice the key point that guarantees the validity of CEM: nonmodal consequents are always evaluated at a single world.

Discussion. Before moving on to consequence, let me notice a feature of the semantics. Path semantics bears an obvious resemblance to Veltman's data semantics ([20]), since it tracks possible trajectories along which an information state might evolve. At the same time, just conditionals are treated very differently. Data semantics uses a strict conditional analysis. In path semantics, conditionals have no quantificational force of their own. The *if*-clause is used

to update the path; the consequent is evaluated at the updated path as any other sentence. Strictly speaking, then, path semantics (somewhat similarly to the restrictor account in [12]) doesn't see conditionals as semantic entities at all. Conditionals are just ordinary sentences prefaced by a path-shifting device, i.e. the *if*-clause.

5 Logical consequence

5.1 Defining consequence

Path semantics allows us to define several notions of consequence. One captures preservation of truth at a path.

Path consequence.
$\phi_1, \ldots, \phi_n \vDash_{\mathrm{P}} \psi$ iff for all paths P such that $[\![\phi_1]\!]^P = 1$, …, $[\![\phi_n]\!]^P = 1$, $[\![\psi]\!]^P = 1$

While path consequence is useful for several purposes, it is not the notion that best captures what follows from a set of accepted premises.[13] But a notion of this sort can be easily defined as follows.

Path-Informational consequence.
$\phi_1, ..., \phi_n \vDash_{PI} \psi$ iff, for all i such that $\phi_1, ..., \phi_n$ are true throughout i, ψ is true throughout i.

Path-Informational consequence is the analog, in the current framework, of Veltman's ([21]) test-to-test validity, or Yalcin's ([22]) informational consequence. Informally, it tracks what follows from an information state that validates certain premises. It is the obvious notion of consequence for assessing consistency and validity for asserted claims in natural language.

Path-Informational consequence vindicates both CEM and IMC, while blocking the collapse of *might*-conditionals onto bare conditionals.[14]

Fact 1. $\vDash_{PI} (\phi > \psi) \vee (\phi > \neg\psi)$

Fact 2. $(\phi > \neg\psi) \wedge (\phi > \Diamond\psi) \vDash_{PI} \bot$

Fact 3. $\phi > \Diamond\psi \nvDash_{PI} \phi > \psi$

Notice also that, despite the validity of IMC, Duality fails.

Duality. $\nvDash_{PI} (\phi > \Diamond\psi) \leftrightarrow (\neg(\phi > \neg\psi))$

6 Conclusion

Path semantics reconciles CEM and the inconsistency of $\phi > \neg\psi$ and $\phi > \Diamond\psi$, solving a problem that in various forms has been discussed since the beginning of modern work on conditionals. My proposal is confined to epistemic conditionals, but the puzzle generalizes to counterfactuals. The results of this paper encourage exploring the prospects for a general semantic framework for conditionals that accommodates both epistemic conditionals and counterfactuals.

[13] To see this, notice that path consequence fails to vindicate one of the signature inference patterns of the semantics in [21] and [22], i.e. what Yalcin calls 'Łukasiewicz's principle'.

Łukasiewicz's principle. $\neg\phi \vDash \neg\Diamond\phi$

[14] Here are some intuitive proofs. As for CEM: $(\phi > \psi) \vee (\phi > \neg\psi)$ is true at all paths, hence it is true throughout any i. As for IMC: if $\phi > \psi$ is true at all paths terminating at i, then i contains no $\phi \wedge \neg\psi$-worlds, hence $\phi > \Diamond\neg\psi$ is false.

References

[1] Fabrizio Cariani and Paolo Santorio. *Will* done better: Selection semantics, future credence, and indeterminacy. forthcoming in *Mind*, 2016.

[2] Alexandre Cremers, Manuel Križ, and Emmanuel Chemla. *Probability Judgments of Gappy Sentences*, pages 111–150. Springer International Publishing, Cham, 2017.

[3] Jonathan St. B. T. Evans and David E. Over. *If*. Oxford University Press, Oxford, 2004.

[4] Kai von Fintel. Bare plurals, bare conditionals, and only. *Journal of Semantics*, 14(1):1–56, 1997.

[5] Kai von Fintel and Sabine Iatridou. If and when if -clauses can restrict quantifiers. unpublished draft, available at http://web.mit.edu/fintel/fintel-iatridou-2002-ifwhen.pdf, 2002.

[6] Anthony S. Gillies. Epistemic conditionals and conditional epistemics. *Noûs*, 38(4):585–616, 2004.

[7] Alan Hájek and N. Hall. The Hypothesis of the Conditional Construal of Conditional Probability. In Ellery Eells, Brian Skyrms, and Ernest W. Adams, editors, *Probability and Conditionals: Belief Revision and Rational Decision*, page 75. Cambridge University Press, 1994.

[8] James Higginbotham. Linguistic theory and davidson's program in semantics. In Ernest LePore, editor, *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, pages 29–48. Cambridge: Blackwell, 1986.

[9] Stefan Kaufmann. Conditioning against the grain. *Journal of Philosophical Logic*, 33(6):583–606, 2004.

[10] Theodore Korzukhin. Dominance conditionals and the Newcomb problem. *Philosophers' Imprint*, 14(9), 2014.

[11] Angelika Kratzer. Partition and revision: The semantics of counterfactuals. *Journal of Philosophical Logic*, 10(2):201–216, 1981.

[12] Angelika Kratzer. *Modals and Conditionals: New and Revised Perspectives*, volume 36. Oxford University Press, 2012.

[13] Manuel Križ. *Aspects of homogeneity in the semantics of natural language*. PhD thesis, University of Vienna, 2015.

[14] David Lewis. Probabilities of conditionals and conditional probabilities. *Philosophical Review*, 85(3):297–315, 1976.

[15] David K. Lewis. *Counterfactuals*. Harvard University Press, Cambridge, MA, 1973.

[16] Clemens Mayr. Alternative questions in a trivalent semantics. forthcoming in *Journal of Semantics*, 2017.

[17] Daniel Rothschild. Do indicative conditionals express propositions? *Noûs*, 47(1):49–68, 2013.

[18] Philippe Schlenker. Conditionals as definite descriptions. *Research on language and computation*, 2(3):417–462, 2004.

[19] Robert Stalnaker. A theory of conditionals. In N. Recher, editor, *Studies in Logical Theory*. Oxford, 1968.

[20] Frank Veltman. *Logic for Conditionals*. PhD thesis, University of Amsterdam, 1985.

[21] Frank Veltman. Defaults in update semantics. *Journal of Philosophical Logic*, 25(3):221–261, 1996.

[22] Seth Yalcin. Epistemic modals. *Mind*, 116(464):983–1026, 2007.

Semantic Abstractionism

Giorgio Sbardolini

The Ohio State University
sbardolini.1@osu.edu

Abstract

Propositions may be defined by an abstraction principle, somewhat along the lines of similar accounts of abstract objects—most famously, the Neo-Fregean account of numbers. I present the basic outlines of such an account, discuss how it compares with existing theories of propositions, and sketch an initial defense from objections. The resulting theory is hyperintensional, improves on alternatives views on questions of granularity, and better fits with Linguistic methodology by codifying the practice of semanticists.

1 The Nature of Propositions

There's some agreement in Philosophy about the role propositions are supposed to play in theories of communication and cognition, but there are different views about what kind of objects play those roles. Typically, it's assumed that propositions are:

1. the semantic values of declarative sentences (relative to a context), and
2. the primary bearers of truth-values, and
3. the objects of propositional attitudes.

The competition is between two main families of accounts:[1] the Possible-World theory, and Structured Propositions theories (of which there are many). To a rough approximation, these contenders differ in giving privilege to job 1 over job 3 (the PW theory), or viceversa (SP theories): the metaphysics of each account is designed for the preferred job.

According to the PW theory ([21, 27]), propositions are sets of possible worlds. The PW theory has by far the best record of successful applications in Semantics, and thus there are good reasons to accept it. However, it has long been known to yield undesirable consequences, and pressure to revise or reject the PW theory is on the rise. Perhaps the most famous undesirable consequence is that, on the PW theory, all necessarily equivalent propositions are identical. I use angle brackets '$\langle ... \rangle$' as a term-forming operator for propositions. So '$\langle$*1021 is prime*$\rangle$' is a term denoting the proposition that 1021 is prime. According to the PW theory, $\langle$*1021 is prime*$\rangle$ and $\langle$*Triangles have 3 sides*$\rangle$ are identical, because both are true in all possible worlds.

However, it appears that such identity conditions are too coarse grained, since it seems that a rational subject might believe that triangles have 3 sides and fail to believe that 1021 is prime, or viceversa. If so, then these should be distinct propositions. The Problem of Necessary Equivalents is a failure to distinguish distinct propositions with the same modal profile.[2]

Different strategies have been explored to account for this problem within the PW theory: one might tinker with the definition of a possible world, or try Stalnaker's two-dimensionalist

[1] There are more options than the ones I list: some are discussed in e.g. [2, 15]. Some accounts reject the characterization I just gave: on "Radical Pragmatics" accounts, semantics disappears, as it were, squeezed between syntax and a very rich pragmatics. Things that play role 3 still exist, but nothing plays role 1.

[2] The criterion for which propositions should be distinct, if it is possible for a rational subject to believe one while failing to believe the other, seems intuitively plausible. Its long history goes back to Frege's Equipollence criterion for the identity of contents, in [8, p. 197]. For a recent discussion, see [28].

analysis ([27]), or rely on a theory of fragmented beliefs (e.g. [7]). On the other hand, the Problem of Necessary Equivalents seems to frustrate the attempt of the PW theory to yield an adequate account of the objects of attitudes, and so many regard it as the symptom of an underlying misconception. Accordingly, its lesson is taken to be that propositions ought to be individuated hyperintensionally.

SP theories are the most important alternative to date. What's common to these theories is the assumption that a proposition is a structured object whose parts correspond to the semantic values of the syntactic constituents of the sentence that expresses it. There are many different SP theories, e.g. Neo-Russellian accounts, and structured Fregean accounts that include 'modes-of-presentation'. On King's view ([17]), a proposition is composed of objects and properties, and a relation that "glues" them together. On Soames's view ([26]), a subject/predicate proposition is the cognitive act of predicating something of an object. On Hanks's view ([13]), a subject/predicate propositions is a cognitive act of predicating something of an object, together with a cognitive act of referring to it. Although different theories have different virtues and shortcomings,[3] and different accounts discriminate among propositions differently (depending on the treatment of singular terms), a common feature of SP theories is that the identity conditions of propositions depend on the identity of the predicates that contribute to expressing them. For example, consider the following pairs:

(1) a. 3 is greater than 2.
b. 2 is smaller than 3.

(2) a. I met a farmer who was feeding a donkey.
b. I met a farmer$_1$ and he$_1$ was feeding a donkey.

Since *x is greater than 2* and *x is smaller than 3* are distinct predicates expressing distinct properties, and acts of predicating them are distinct cognitive acts, SP theories have the consequence that the propositions expressed by (1a) and (1b) are distinct. Since the proposition expressed by (2b)—where the index '1' indicates the intended reading of the pronoun—contains a conjunction, or the act of conjoining, and the proposition expressed by (2a) doesn't, the propositions expressed by (2a) and (2b) are distinct, according to SP theories. More dramatically, for the same reason, $\langle A \rangle$ and $\langle A \textit{ and } A \rangle$ are regarded as distinct, whatever sentence A is.

A popular complaint is that on SP theories propositions are individuated too finely. Sometimes this complaint is put in terms of a failure to reflect philosophers' intuitions about "saying the same thing": if this is the objection, I believe that King is quite correct in dismissing it ([18]). I also believe, however, that the more interesting objection is methodological: there is no use in semantics for a notion of proposition that distinguishes between pairs such as (1a) and (1b), and (2a) and (2b) respectively. The distinction in propositional content between $\langle A \rangle$ and $\langle A \textit{ and } A \rangle$ is one without a difference, for the behavior of competent speakers of English does not indicate that there's any semantic difference between (1a) and (1b), and (2a) and (2b) respectively, and so if one insists that these all express different propositions, important semantic generalizations are missed. Problems of Fineness of Grain are failures to identify the same proposition as being expressed by different sentences. According to some critics, this problem indicates that SP theories fail to characterize a notion of proposition that plays the role of contents of declarative sentences (role 1), as characterized by truth-conditional semantics.

I shall now outline a third option, and try to convince you that it's better than both the PW and SP theories. The view of propositions I develop here is designed to codify the methodology of empirical research in semantic theory.

[3] A shortcoming specific to Soames's view is that there are no propositions, as he defines them, expressed by true negative existential sentences. See Soames's discussion in [26, p. 230].

2 Abstraction Principles

An abstraction principle is a statement of the form:

$$\forall\alpha,\beta(\$\alpha = \$\beta \equiv \alpha \sim \beta)$$

where '$' is a term-forming operator (like '$\langle ... \rangle$'), and '$\sim$' is an equivalence relation between objects in the domain of the quantifiers. The most famous abstractionist account is the Neo-Fregean account of numbers ([32, 12]). According to Neo-Fregeans, numbers are defined by Hume's Principle:

$$\text{(HP)} \qquad \forall F,G(Nx : Fx = Nx : Gx \equiv \Theta(F,G))$$

where 'F' and 'G' are variables ranging over concepts, '$Nx : \Phi x$' is the term-forming operator *The number of*, and 'Θ' abbreviates a second order statement of equinumerosity. So HP says, intuitively, that the number of the concept F is identical to the number of the concept G iff F and G are equinumerous. There are other abstraction principles, such as Frege's principle for directions (in *Grundlagen*, section 65). Thanks to the Neo-Fregean program, the logic and philosophy of abstraction principles have been extensively discussed in recent years, and are now fairly well understood.

Now consider the following inspiring passage from Frege's *Begriffschrift* (section 3):

> I remark that the contents of two judgments may differ in two ways: either the consequences derivable from the first, when it is combined with certain other judgments, always follow also from the second, when it is combined with these same judgments, [and conversely,] or this is not the case. ... I call that part of the content that is the *same* in both, the *conceptual content*.

Related passages may be found throughout Frege's work, but let's set aside questions of interpretation.[4] Perhaps an abstraction principle for propositions ('conceptual contents') can be extracted from this passage. Frege seems to suggest that given two 'judgments' A and B, the proposition expressed by A is identical to the proposition expressed by B iff A and B have the same consequences given the same assumptions. So, an abstraction principle for propositions might be the following ('Semantic Abstraction'):

$$\text{(SA)} \qquad \forall A,B(\langle A\rangle = \langle B\rangle \equiv \forall\Gamma\forall C(\Gamma, A \text{ entail } C \equiv \Gamma, B \text{ entail } C))$$

i.e. the proposition that A is identical to the proposition that B iff anything C entailed by A together with Γ is entailed by B together with Γ. I take Γ to be a possibly empty set of things of the same type as things in the range of A, B, and C. SA identifies the proposition expressed by A and B just in case A and B are, as it were, "equientailing": they entail the same things under the same assumptions. Semantic Abstractionism is the view that propositions are the objects defined by SA.[5] I shall now discuss what the variables in SA range over, and discuss the notion of entailment. This will take up most of the paper. In the last section, I will briefly address some features of Semantic Abstractionism that in some way or other apply to all abstractionist projects, including a few remarks on the Caesar Problem.

[4] See [10, p. 188], [9, p. 70], and the Equipollence criterion referred to in fn. 2.

[5] The idea of defining propositions by abstraction is not entirely new, and not just because it may have been suggested by Frege. An attempt at such a definition is made by Hale in [12, p. 91–116], though the background and goals of Hale's discussion are completely different from mine. Another account is discussed and negatively assessed by Wrigley in [33]. In part, Wrigley worries about what types of objects the variables A and B range over. I shall address this point below. Remaining worries raised by Wrigley are dealt with in [30].

3 Natural Language Syntax

I shall assume, as in informal explanation of the concept of proposition underlying SA, that propositions are things expressed by sentences. In my opinion, SA is best understood as abstracting from (something like) natural language sentences—whether we may define propositions abstracting from sentences of some formal language is not a simple question, and I set it aside.

We should be clear on the notion of sentence that is relevant here. Since it's implausible to take propositions to be defined by sentences regardless of context (for some sentences express different propositions in different contexts), the variables in SA should range over utterances, i.e. pairs of a sentence and a context.[6] However, utterances are sometimes understood as concrete physical objects: actually occurring sequences of sounds. But it's implausible to assume that all propositions are expressed by actually occurring sequences of sounds (there aren't enough of them). So I shall take the variables in SA to range over *utterance-types*, i.e. pairs of a sentence-type and a context. Generalizing beyond the concrete physical objects that provide evidence for a theory is a straightforward generalization made for scientific purposes. Perhaps not unproblematic, but very common.

The variables in SA range over utterance-types of declarative sentences of a natural language. Two related questions may arise. Which language are we talking about? And also: How are sentence-types identified? I will address the first question at the end of this section. In the answer to the second question, it is crucial that we distinguish different sentences without relying on the assumption that they express different propositions: that would be circular. Abstraction principles, like SA, succeed in establishing the identity conditions of the objects denoted by singular terms on the left-hand side, only if the conceptual resources employed on the right-hand side do not already presuppose what the principle itself should establish. If there is no explanatory priority of the right-hand side on the left-hand side of the main biconditional in SA, the attempted definition by abstraction fails.

To avoid this potential problem, I shall identify sentence-types syntactically. There is a tendency in Philosophy to think of sentences as phonologically individuated objects, but, in my opinion, this may be no more than an old empiricist prejudice. Better to abstract away from the phonological description (Phonological Form) of sentences, and identify them by their syntactic description. You can think of syntactically individuated sentence-types, in the tradition of Chomsky, as those objects recursively generated by the grammar that get "sent off" to the 'conceptual-intentional interface' for semantic interpretation—sometimes these are called 'Logical Forms', or 'Sentences in the Language of Thought'. For present purposes, it is unnecessary to think of sentence-types as mental objects, though we might. But for all I say here, they might be mathematical objects of some kind (see fn. 7).

Semantic Abstractionism is plausible only if the background syntactic theory distinguishes between sentence-types that, relative to the same context, express different propositions. Thus to some degree, the truth of SA depends on details of syntactic theory, but that seems fair. Indeed, the truth of SA depends on the assumption, which is commonly made by different theories of grammar, that semantic composition and syntactic dependence work in parallel.[7] Semantic

[6] By a context I shall take, for simplicity, a Lewisian centered world ([20]): it seems reasonable to think, at least to a first approximation, that all the information required for the resolution of context-sensitivity, broadly understood, is packed into a centered world. A more sophisticated option, perhaps, is to take a context to be a particular set of centered worlds. The choice bears on accounts of context-sensitivity, and deserves more discussion than a footnote, but these and possibly other options are compatible with the present account.

[7] There are many theories of grammar, but my claim holds of at least Generative Grammar (e.g. [3]) and Head-driven Phrase Structure Grammar ([23]). These differ in many details, one of which is the metaphysical status of the objects they study (which I hinted at in the previous paragraph). A major difference between GG and HPSG is the analysis of syntactic dependence (movement vs. structural identity). There are several other

Abstractionism can thus afford to be largely neutral on the choice of background grammar, since it is uncontroversial that the syntax has the resources to make some indispensable distinctions, such as the ones below. Consider (the source of examples (3) and (5) is [1]):

(3) Teacher strikes idle kids.

This sentence is structurally ambiguous: (3) could mean that kids have become idle as a consequence of teacher strikes, or that a teacher hit some lazy kids. In order to account for structural ambiguity, the most indispensable item in the grammar is some notion of phrase structure, which is part of an account of syntactic dependence. Phrase structure analysis is quite fundamental to syntactic theory—but somewhat trivial for the syntax of artificial languages because of the small set of syntactic categories.[8] On a rough approximation, there are two phrase structure analyses of (3):

(4) a. $[_{S}\ [_{NP}$ Teacher strikes $]\ [_{VP}$ idle kids $]\ ]$
b. $[_{S}\ [_{NP}$ Teacher $]\ [_{VP}$ strikes idle kids $]\ ]$

In the former, *Teacher strikes* is the NP to the left of the verb, with the head noun *strikes*, and *idle kids* is the VP; in the latter, *Teacher* is the left NP, and *strikes idle kids* is the VP. So there are (at least) two sentence-types corresponding to (3). Lexical ambiguity works otherwise:[9]

(5) Doctor testifies in horse suit.

(5) could mean that a doctor testified in a legal case involving a horse, or that a doctor testified dressed like a horse. As customary, lexical ambiguity is eliminated by distinguishing between expressions. There's two words *suit* in English, i.e. distinct lexical entries with the same phonological description. That these are distinct words can be established independently of any semantic knowledge of the kind that might threaten the claim of explanatory asymmetry regarding the two sides of SA. The key notion in this case is cross-linguistic comparability. Since the two words *suit*, for example, are systematically distinguished even in languages closely related to English, these are different words, with a different history.[10] The standard assumption in Linguistics is that the lexicon is the source of arbitrariness in language: if there was a grammatical explanation for why English apparently correlates tuxedos and court trials, the correlation should be found across languages.[11]

World knowledge may be indispensable for speakers to arrive at the preferred interpretation in some cases, but that doesn't matter. What's crucial for an account of propositions by abstraction is that utterance-types expressing different propositions in the same context can be systematically distinguished without relying on the identity of the propositions they express. This is indeed the case.

At the beginning of this section, I raised another question: to which language do these sentence-types belong? Which language-specific grammar is the one whose sentences we abstract propositions from? The answer is: it doesn't matter too much. The extent to which

differences, both conceptual and empirical: see the Introduction to [23].

[8] According to [4, p. 86], the notion of phrase structure goes back to Arnauld and Lancelot's *Port Royal Grammar* (1662).

[9] There are other kinds of ambiguity. Contextual ambiguity (e.g. about the antecedent of unbound anaphors) is taken care of by building contexts into the notion of an utterance-type, and phonological ambiguity (e.g. *right* vs. *rite*) is irrelevant here. Scope ambiguity is another, rather different, case of structural ambiguity, but it's handled similarly. Finally, type-shifting principles may introduce some kind of "ambiguity", but insofar as they apply to resolve syntax-semantics mismatches, they should pose no more problem than structural ambiguities.

[10] It helps to think of the metaphysics of words in the way [16] recommends we do.

[11] For example, Kratzer relies on this standard assumption in her argument that modal verbs are not ambiguous, in "What 'Must' and 'Can' Must and Can Mean", now chapter 1 of [19].

it matters is that, to avoid needless complication, sentence-types should belong to a single grammar. The question of which grammar is then answered by the following assumption:

> *Effability*
> Every proposition (if it can actually be expressed at all) can be expressed in every natural language.

Let's set aside the question whether there are actually inexpressible propositions. Any proposition that *can* actually be expressed, can be expressed in every natural language. Of course, languages are going to do it in different ways, depending on what's in their lexicon, on what has to be conventionally conveyed, on what can be backgrounded (e.g. by presupposing), and on what's socially acceptable for a speaker to say. Effability is a strong empirical hypothesis, and an idealization about the availability of lexical resources, but it's a widespread working assumption in empirical semantic research, and I shall not challenge it.[12]

4 Natural Language Entailments

I shall now discuss the notion of entailment in SA. It is crucial to distinguish between *natural language entailment* and *formal entailment*. The relevant notion for SA is the former. Formal entailment, i.e. necessary truth preservation, is well-understood. Clearly though, if the relevant notion for SA was formal entailment, then SA wouldn't distinguish among necessarily equivalent sentences, and no progress would be made on the Problem of Necessary Equivalents. For brevity, I shall use 'entailment' for the natural language notion from here on.

Entailment is a relation between a set of utterance-types and an utterance-type. It's therefore context-sensitive, and moreover it should relate utterance-types of different contexts, otherwise we couldn't identify propositions expressed in different contexts—see [24] for a discussion of a formal definition of cross-contextual validity. Furthermore, it seems reasonable to assume that necessary truth preservation is a necessary condition on entailment, so that whenever A entails B, it follows that, necessarily, A is true only if B is true. But it is not a sufficient condition. So necessary truth preservation can be a heuristic guide to entailment, and indeed of crucial importance, because we have sophisticated mathematical techniques to study necessary truth preservation. Entailment is, more generally, a grammatical relation, that can be studied by generalizing probabilistically over competent speakers' patterns of judgment in an experimental setting. Its theory is partly a matter to be settled empirically, and partly by global theoretical considerations.

Semantic Abstractionism is a view of the nature of propositions designed for the methodology of Linguistics. Typically, the study of the content of linguistic expressions is carried out by means of judgment elicitation tasks, during which a speaker may be tested as to whether she takes a given sentence to be entailed in context. What this methodology seems to indicate is that information about contents is provided primarily by the investigation of what entailments obtain in a context. An abstractionist metaphysics of propositions reflects the epistemological underpinnings of linguistic methodology. The resulting view is conservative also in the sense that, once propositions are defined, intensions are still available, and each proposition can be assigned a function from worlds to truth-values (worlds may already be needed, as contexts: see fn. 6).

It seems to me that the existence of patterns of entailment is something of a pre-condition on the plausibility of semantics as an empirical discipline, and there is no reason to be skeptical

[12] See the favorable discussion in [31]. I also rely on their assessment of the persistent myth that native speakers of different languages *must have* different cognitive abilities, like grasping different propositions.

about it. No one doubts that data from speakers acceptance and rejection responses fall into general patterns, unless one is skeptical about semantics itself.[13] Notice that SA does not require that the notion of entailment be an equivalence relation (I do assume, below, that entailment is reflexive; but this doesn't seem particularly problematic). Abstraction does require the definition of an equivalence relation between utterance-types, but such relation is: *the set of things entailed by x and* Γ *has the same members as the sets of things entailed y and* Γ. Assuming entailments data are robust enough for the development of semantics, I think it's plausible to say that such relation is well-defined, and then it is certainly an equivalence relation.[14]

It is important to emphasize that generalizations about entailments are probabilistic, as we are trying to characterize a relation that depends on the grammar: regularities of speakers' linguistic competence, generalizing away from performance limitations. A first consequence of this point is that, should a competent speaker accept one of a pair of utterance-types that express the same proposition (according to the population which she belongs to), and fail to accept the other, that would indicate a performance error (of which there are various kinds). A second consequence is that no claim of analiticity should be attached to SA—unlike (perhaps) HP, the epistemology of SA is that of an empirical generalization.[15]

Not everything is an entailment: for instance, Gricean conversational implicatures are not entailments. There are standard tests to determine what speakers take to be entailed, and what they derive instead from world knowledge and expectations about other people's behavior. Indeed, implicatures are not considered part of the semantic content of a sentence. Therefore, part of the question what counts as an entailment bears on theoretical choices about what counts as semantics. None of this warrants skepticism about the notion of entailment.

We don't know much about the global shape of a theory of natural language entailment (yet). It's not going to look like any of the formal systems logicians have studied, although, locally (i.e. relative to some parameter), it might. Entailments are sensitive to various features of communication, well beyond the usual parameters of context-sensitivity. It's possible, but consistent with SA, that a theory of entailment will not be axiomatizable, but that of course doesn't mean that there are no generalizations to be made. Perhaps the best approach should be some kind of pluralism about entailment. This is not necessarily a problem.

A different worry is that the notion of linguistic competence, that I rely upon in an explanation of what counts as an entailment, might be where semantic notions are illegitimately sneaking in. Of course, if judgments about entailments are not constrained by competence, all bets are off. But competence is sometimes understood as involving some kind of privileged epistemic status on part of speakers, perhaps some kind of a priori knowledge about the language. Maybe linguistic competence consists in part in the ability to recognize how many propositions are expressed by two utterance-types. If that's the case, the required explanatory asymmetry between the two sides of SA is flouted.

[13] In which case there would be little need for a *semantic* notion of content—see fn. 1. Anyway, such skepticism is not to be confused with, for instance, Glanzberg's skepticism (in [11]) that anything like *Tarski's notion of logical consequence* can be reconstructed from natural language. Glanzberg agrees, of course, that there are semantic regularities in natural language.

[14] So, SA does not require that to establish how many propositions are expressed by two utterance-types, one should merely check for mutual entailments. The semantics is, of course, still compositional, and generalizations about contents are the result of testing speakers on a variety of related tasks—consistently with the methodology that justifies SA. Semantic Abstractionism, therefore, does not come with an account of Frege's Puzzle, but it was never meant to.

[15] One issue is that a grasp of the left-hand side of SA by a speaker does not provide a priori justification for reference to propositions: whether SA holds is an a posteriori matter. There is also a deeper and related issue. In my discussion, I do rely on a (partial) explanation of the concept of a proposition that doesn't follow from SA: namely, that propositions are things expressed by sentences. So I don't claim that grasp of SA alone *suffices* for reference. This is a radical departure from the Neo-Fregean conception of abstraction.

To defuse the circularity worry, I shall sketch a probabilistic account of competence:

Competence
Competence is better than Random.

The key insight behind this slogan[16] is to treat competence as the probability to give the right response, and to stipulate that competent speakers are those who perform better than a random function (by a statistically significant measure) at least in tasks of production and recognition. Consider a subject S and a fair coin. To establish whether S is competent in L, we assign production and recognition tasks. For instance, in a recognition task, S is shown sequences of lexical items of L and is asked for each sequence whether it is a sentence of L. In another room, a fair coin is flipped for each sequence shown to S, and the coin "says" that a sequence is a sentence of L just in case it lands Heads. So the coin's "responses" to the task are random. S is competent only if she scores significantly better than the coin in the long run. Production tasks are designed in a similar fashion. The result is a somewhat minimal condition on competence, a condition of adequacy that is perhaps necessary but insufficient—and so, the probabilistic account I sketched is not a conceptual analysis. Much more could be said, but unlike epistemological accounts of competence, the probabilistic account seems to promise a strategy to defuse circularity worries about SA.

5 Ontology and Troubles

Many questions about the ontology of abstraction principles have been already addressed by existing literature. Semantic Abstractionism is compatible with different ways of understanding the ontology and metaontology of abstraction.[17] Following [22], abstractionist projects can be understood as implementing a form of *metaontological minimalism*, since objects defined by abstraction are ontologically "thin": very little is required on the world for their existence. On the other hand, precisely because the demands on reality are little, abstractionist views tend to support forms of *ontological maximalism* ([6]): there's a lot of objects of the kind so defined. Further questions can be addressed within the basic framework for the metaphysics of abstraction developed by existing literature.

Semantic Abstractionism inherits some potential problems of any abstractionist accounts. The interesting question for present purposes is whether the present perspective adds anything illuminating to the debates about these well-investigated difficulties, and there might be something to be said about the Caesar Problem.[18] What the Caesar Objection is has been clarified in recent years, and for my brief remarks here I rely mostly upon [14]. The difficulty for Frege's

[16] This is inspired by the account of competence in decision-making of Condorcet's *Essai* on voting systems (1785).

[17] According to e.g. [25], abstraction principles yield a *platonistic* account of abstract objects, at least in the sense that objects defined by abstraction exist necessarily. However, there's reason to doubt that all propositions exist necessarily, the reason being a common view of singular propositions. There are ways to account for contingently existing propositions that are compatible with SA. A strategy might be to consider whether sentence-types exist necessarily. Perhaps they don't, since their identity conditions depend on the linguistic resources available in actual human languages, especially their lexicon, and it seems that there might be merely possible languages just like ours but whose speakers lack the resources to refer to actual individuals with whom they don't have any cognitive contact. In these non-actual languages, there will be no propositions about actual individuals. So, whether contingently existing propositions defined by SA might be accommodated with some approaches to the metaphysics of abstraction deserves careful consideration.

[18] The *other* major difficulty facing any abstractionist account is the Bad Company Objection. This is a challenge to explain the difference between acceptable abstraction principles and unacceptable ones. I don't think that the present project has much new to contribute to this. For a recent discussion and overview, see [5].

project can be described as follows. The job of HP is to establish the identity conditions of numbers. This is done by establishing the truth-conditions of formulas in which the identity sign is flanked by singular terms of a certain form that denote numbers. However, in order to evaluate the truth of e.g. '$\exists x(x = Ny : Fy)$', i.e. something is the number of F, we should be able to evaluate '$x = Ny : Fy$' for every value of 'x' in the domain of the quantifier. Given Frege's universalist conception of logic, *every* object falls in the domain of quantifiers that range over numbers, including Julius Caesar. Indeed, it is crucial for Frege's conception of numbers as objects that numbers be of the same logical type as any other object. But while HP establishes the truth-conditions of identity statements about numbers where both singular terms are of the form '$Nx : \Phi x$', HP is no help on the question whether Caesar is the number of F. Of course he is not, but that's not thanks to HP—just like Frege said.

Perhaps we can go again through the steps above, replacing 'propositions' for 'numbers' everywhere. Thus SA establishes the truth-conditions of identity statements in which the identity sign is flanked by singular terms of a certain form that denote propositions, and an objectual understanding of quantification over propositions requires the evaluation of formulas '$x = \langle A \rangle$' for every object in the domain. As I pointed out, the Caesar Problem hits Frege with particular force given his universalist conception of logic. But there may be reasons to think that quantifiers over intensional entities, such as propositions, may be subject to special restrictions that don't apply everywhere. Such reasons may come from an analysis of the intensional paradoxes (which I shall not discuss here, see [29]), and may be further justified by the homely thought that the laws of semantics are the laws of a special science. This thought is available also to a Fregean universalist. So we may suppose that the domain of the quantifiers in SA only include intensional objects of the right type. This restriction helps us dodge the most serious consequences of the Caesar Problem. Further questions, no doubt, still remain.

6 Conclusions

I have discussed and clarified various aspects of SA, an abstraction principle for propositions. The rough sketch given here suffices, I hope, to show some of the virtues of Semantic Abstractionism. Presumably, competent speakers of English won't regard *1021 is prime* and *Triangles have 3 sides* as entailing one another to begin with, though each entails itself. So $\langle$*1021 is prime*$\rangle$ and $\langle$*Triangles have 3 sides*$\rangle$ are not identical, and it's consistent to accept one and fail to accept the other. Moreover, it seems plausible to say that there's no context in which I met a farmer who was feeding a donkey such that a competent speaker would not accept *I met a farmer*$_1$ *and he*$_1$ *was feeding a donkey* as entailed, and vice-versa. This kind of considerations indicates that $\langle$*I met a farmer who was feeding a donkey*$\rangle$ and $\langle$*I met a farmer*$_1$ *and he*$_1$ *was feeding a donkey*$\rangle$ are identical. So, it seems, an account of propositions based on SA scores higher than the PW theory on the Problem of Necessary Equivalents, and higher than SP theories on Fineness of Grain. The identity conditions of propositions fixed by SA are determined neither by the identity conditions of sets of possible worlds, as for the PW theory, nor by the identity of predicates or acts of predication, as for SP theories. Rather, propositions are identified in part syntactically, by the identity of sentence-types that express them relative to a context, and in part cognitively, by what speakers take to be entailed by an utterance in a context.

References

[1] C. Bucaria. Lexical and syntactic ambiguity as a source of humor: the case of newspaper headlines. *Humor*, 17(3):279–309, 2004.

[2] R. Carston. *Thoughts and Utterances*. Oxford: Blackwell, 2002.

[3] N. Chomsky. *The Minimalist Program*. Cambridge, MA: MIT Press, 1995.

[4] N. Chomsky. *Cartesian Linguistics*. Cambridge: Cambridge University Press, 3rd edition, 2009.

[5] R. Cook. Conservativeness, Cardinality, and Bad Company. In Ebert and Rossberg, editors, *Abstractionism*, pages 223–246. Oxford: OUP, 2016.

[6] M. Eklund. Neo-fregean ontology. *Philosophical Perspectives*, 20(1):95–121, 2006.

[7] A. Elga and A. Rayo. Fragmentation and information access, ms.

[8] G. Frege. A Brief Survey of My Logical Doctrines. In Hermes, Kambartel, and Kaulbach, editors, *Posthumous Writings*, pages 197–202. Oxford: Blackwell, 1979.

[9] G. Frege. Letter to Husserl 9/12/1906. In Hermes, Kambartel, Kaulbach, Tiel, and Veraart, editors, *Philosophical and Mathematical Correspondence*, pages 70–71. Oxford: Blackwell, 1980.

[10] G. Frege. On Concept and Object. In McGuinness, editor, *Collected Papers on Mathematics, Logic, and Philosophy*, pages 182–194. Oxford: Blackwell, 1984.

[11] M. Glanzberg. Logical Consequence and Natural Language. In Caret and Hjortland, editors, *Foundations of Logical Consequence*, pages 71–120. Oxford: OUP, 2015.

[12] B. Hale and C. Wright. *The Reason's Proper Study*. Oxford: Clarendon Press, 2001.

[13] P. Hanks. *Propositional Content*. Oxford: OUP, 2015.

[14] R. Heck. *Frege's Theorem*. Oxford: OUP, 2011.

[15] D. Hunter and G. Rattan. *New Essays on the Nature of Propositions*. New York: Routledge, 2015.

[16] D. Kaplan. Words. *Proceedings of the Aristotelian Society*, 64:93–119, 1990.

[17] J. King. *The Nature and Structure of Content*. Oxford: OUP, 2007.

[18] J. King. On fineness of grain. *Philosophical Studies*, 163:763–781, 2013.

[19] A. Kratzer. *What 'Must' and 'Can' Must and Can Mean*. Oxford: OUP, 2012.

[20] D. Lewis. Attitudes de dicto and de se. *The Philosophical Review*, 88(4):513–543, 1979.

[21] D. Lewis. *On the Plurality of Worlds*. Oxford: Blackwell, 1986.

[22] Ø. Linnebo. Metaontological minimalism. *Philosophy Compass*, 7(2):139–151, 2012.

[23] C. Pollard and I. Sag. *Head-driven Phrase Structure Grammar*. Chicago: The University of Chicago Press, 1994.

[24] A. Radulescu. The logic of indexicals. *Synthese*, 192(6):1839–1860, 2015.

[25] A. Rayo. *The Construction of Logical Space*. Oxford: OUP, 2013.

[26] S. Soames. *Rethinking Language, Mind, and Meaning*. Princeton, NJ: Princeton University Press, 2015.

[27] R. Stalnaker. *Inquiry*. Cambridge, MA: MIT Press, 1984.

[28] M. Textor. Frege's recognition criterion for thoughts and its problems. *Synthese*, pages 1–20, 2017.

[29] D. Tucker and R. Thomason. Paradoxes of intensionality. *The Review of Symbolic Logic*, 4(3):394–411, 2011.

[30] M. Vignolo. Abstracting propositions: How many of them do we need? *Conceptus*, 39(95):7–176, 2010.

[31] K. von Fintel and L. Matthewson. Universals in semantics. *The Linguistics Review*, 25:139–201, 2008.

[32] C. Wright. *Frege's Conception of Numbers as Objects*. Aberdeen: Aberdeen University Press, 1983.

[33] A. Wrigley. Abstracting propositions. *Synthese*, 151(2):157–176, 2006.

On question exhaustivity and NPI licensing*

Bernhard Schwarz

McGill University, Montreal, Quebec, Canada
bernhard.schwarz@mcgill.ca

Abstract

Guerzoni and Sharvit (2007) discovered that the licensing of weak negative polarity items (NPIs) by embedded questions requires a (strongly) exhaustive interpretation in the sense of Groenendijk and Stokhof (1982). This paper points out that under the view that wh-questions themselves are ambiguous between exhaustive and non-exhaustive meanings (George 2011, Guerzoni and Sharvit 2014, Nicolae 2015, Theiler et al. 2016), Guerzoni and Sharvit's observation falls out from an analysis of NPI licensing by questions developed in Krifka (1995) and van Rooy (2003), an analysis that construes questions' strength as their information theoretic entropy (Shannon 1948). Restrictions on NPIs in disjunctive questions (Schwarz in press) and singular *which*-questions are presented as further support for the account.

1 Introduction

Guerzoni and Sharvit (2007) propose that the licensing of weak Negative Polarity Items (NPIs) like *any* or *ever* by embedded questions requires (strong) exhaustivity in the sense of Groenendijk and Stokhof (1982). That is, they propose that an embedded question licenses an NPI only if the semantics of the embedding predicate makes reference to (strongly) exhaustive answers. As evidence for this claim, call it the *exhaustivity-licensing generalization*, Guerzoni and Sharvit point to the pattern illustrated in (1) and (2), which suggests that wh-questions function as NPI licensors when embedded under *wonder* or *know*, but not *surprise*.

(1) Dan wonders who said anything.

(2) a. Dan knows who said anything.
b. *It surprised Dan who said anything.

The exhaustivity-licensing generalization is supported by the contrast in (2) in virtue of truth conditional evidence discovered by Heim (1994) indicating that *know* but not *surprise* makes reference to the negations of so-called Hamblin answers, hence that *know* but not *surprise* makes reference to exhaustive answers to the embedded question. Fleshing out their case for the exhaustivity-licensing generalization, Guerzoni and Sharvit argue that *wonder*, like *know* and unlike *surprise*, makes reference to exhaustive answers.

Under one current approach to (non-)exhaustivity in the interpretation of embedded questions, it is wh-questions themselves that are semantically ambiguous between exhaustive and non-exhaustive interpretations (George 2011, Guerzoni and Sharvit 2014, Nicolae 2015, Theiler et al. 2016). On this view, a question embedding structure is intuited to make reference to (non-)exhaustive answers in virtue of the embedding predicate selecting for the (non-)exhaustive

*For discussion related to this project, thanks to Luis Alonso-Ovalle, Dan Goodhue, Aron Hirsch, Tim O'Donnell, Junko Shimoyama, and Michael Wagner, as well as the (other) members of the McGill Semantics Research Group. Thanks to Brian Buccola, Dan Goodhue, and Aron Hirsch for providing English judgments. This research was supported by the Social Sciences and Humanities Research Council (SSHRC), grants #435-2016-1448 and #435-2013-0592.

reading of the question. The exhaustivity-licensing generalization then holds in virtue of questions being able to license NPIs only if they are exhaustive.

The task is then to account for the latter condition, hereafter the *exhaustivity-licensing condition*. The main objective is to demonstrate that the exhaustivity-licensing condition can be made to fall out as an immediate consequence of an analysis of NPIs in questions developed in Krifka (1995) and van Rooy (2003). This analysis that refers to question's information theoretic entropy (Shannon 1948), and below is dubbed *strength-as-entropy analysis*.

Before proceeding, note that analyzing exhaustivity and non-exhaustivity as properties of question meanings themselves leads to the prediction that, barring any semantic or pragmatic factors that obviate exhaustivity, unembedded questions too can be NPI licensers. As is well known (Klima 1964), this prediction is correct, borne out by examples like (3), where the complement clause featured in (1) and (2) appears as a matrix question.

(3) Who said anything?

After outlining a classic baseline analysis of NPI licensing in terms of strength reversal (Ladusaw 1979, Kadmon and Landman 1993) in section 2, and its elaboration for the case of questions under the strength-as-entropy analysis in section 3, section 4 shows how this account derives the exhaustivity licensing generalization. Section 5 offers additional support for this account from disjunctive questions and singular *which*-questions.

2 NPI licensing under strength reversal

The account of Krifka (1995) and van Rooy (2003) assumes a baseline theory of NPI licensing of the sort pioneered in Kadmon and Landman (1993), which builds on Ladusaw (1979) and is further developed in Krifka (1995), Lahiri (1998), and Chierchia (2013). In a current elaboration, the theory assumes that, in addition to the actual denotation $[\![\,\phi\,]\!]$, grammar assigns a linguistic expression ϕ a set of alternative semantic values $[\![\,\phi\,]\!]^{\mathrm{ALT}}$. Alternative sets feed into the theory of NPIs and their licensing as outlined in (4).

(4) a. $\forall$f' [f' $\in [\![\,\mathrm{NPI}\,]\!]^{\mathrm{ALT}} \rightarrow$ f' $\subset [\![\,\mathrm{NPI}\,]\!]$] NPI semantics
b. $\forall$f' [f' $\in [\![\,\ldots\mathrm{NPI}\ldots\,]\!]^{\mathrm{ALT}} \rightarrow [\![\,\ldots\mathrm{NPI}\ldots\,]\!] \subset$ f'] NPI condition

According to the *NPI semantics* in (4a), the actual denotation of a NPI is strictly weaker than any of the alternatives, while the *NPI condition* in (4b) requires that in some larger syntactic domain *[…NPI …]*, this strength relation be reversed, with the domain's actual denotation being strictly stronger than each of the alternatives.

As shown in (5), alternatives for NPI *anything* are taken to be existential quantifiers that differ from *anything*'s actual denotation in that they have a narrower domain. This ensures that, as stated in (6), each alternative value strictly entails (in a generalized sense) the actual denotation, thereby instantiating the NPI semantics in (4a).

(5) a. $[\![\,\mathrm{anything}\,]\!] = \lambda\mathrm{P}.\lambda\mathrm{w}.\ \exists\mathrm{x}{\in}\mathrm{D}\ [\mathrm{P}(\mathrm{x})(\mathrm{w})]$
b. $[\![\,\mathrm{anything}\,]\!]^{\mathrm{ALT}} = \{\lambda\mathrm{P}.\lambda\mathrm{w}.\ \exists\mathrm{x}{\in}\mathrm{D}'\ [\mathrm{P}(\mathrm{x})(\mathrm{w})]\colon \mathrm{D}'{\subset}\mathrm{D}\}$

(6) $\forall$f' [f' $\in [\![\,\mathrm{anything}\,]\!]^{\mathrm{ALT}} \rightarrow$ f' $\subset [\![\,\mathrm{anything}\,]\!]$]

Alternatives are assumed to expand through point-wise composition (Hamblin 1973, Rooth 1985), yielding sets of alternative properties for a verb phrase like *said anything*, as in (7).

As recorded in (8), in this syntactic domain, the direction of entailment between the actual denotation and the alternatives is of course preserved, rather than reversed.

(7) a. $[\![$ said anything $]\!] = \lambda y.\lambda w.\ \exists x \in D$ [y saw x in w]
b. $[\![$ said anything $]\!]^{ALT} = \{\lambda y.\lambda w.\ \exists x \in D'$ [y saw x in w]: $D' \subset D\}$

(8) $\forall f'$ [$f' \in [\![$ said anything $]\!]^{ALT} \rightarrow f' \subset [\![$ said anything $]\!]$]

How, then, is the NPI condition (4b) met in a question like (3)? The answer given in Kadmon and Landman (1990), Krifka (1995), and van Rooy (2003) is that the NPI condition can be satisfied by the question as a whole. That is, as stated in (9), the proposal is that, under a suitable notion of strength, the actual question denotation can be stronger than each of the alternatives.

(9) $\forall Q'$ [$Q' \in [\![$ who said anything? $]\!]^{ALT} \rightarrow [\![$ who said anything? $]\!] \subset Q'$]

Naturally, this answer requires a proper construal of the strength relation $\subset$, one that is applicable to questions. The construal of the strength relation, in turn, is dependant on the semantic analysis of questions. These issues are addressed in the next section.

3 Question strength as entropy

The notion of question strength proposed in van Rooy (2003) builds on the classic question semantics of Groenendijk and Stokhof (1982). Groenendijk and Stokhof posit the (strongly) exhaustive notion of question meaning referred to in section 1. Such a question meaning determines a set of propositions that partitions the set of possible worlds. For concreteness, suppose the domain contains just two individuals, a and b. If **S** is the property denoted by *said anything*, the *Hamblin answers* to (3) are then the propositions **S**(a) and **S**(b). The denotation of (3) is the partition shown in (10a), whose cells can be obtained by conjoining the two Hamblin answers and their negations. Pointwise composition yields the set of alternative question meanings shown in (10b).

(10) a. $[\![$ who said anything? $]\!] =$
$\{\mathbf{S}(a) \cap \mathbf{S}(b), \neg\mathbf{S}(a) \cap \mathbf{S}(b), \mathbf{S}(a) \cap \neg\mathbf{S}(b), \neg\mathbf{S}(a) \cap \neg\mathbf{S}(b)\}$
b. $[\![$ who said anything? $]\!]^{ALT} =$
$\{\ \{\mathbf{S'}(a) \cap \mathbf{S'}(b), \neg\mathbf{S'}(a) \cap \mathbf{S'}(b), \mathbf{S'}(a) \cap \neg\mathbf{S'}(b), \neg\mathbf{S'}(a) \cap \neg\mathbf{S'}(b)\}:$
$\mathbf{S'} \in [\![$ said anything $]\!]^{ALT}\}$

Assuming this question semantics, van Rooy (2003), generalizing a proposal in Kadmon and Landman (1990) and developing suggestions in Krifka (1995), defines the strength relation between question meanings Q and Q' as in (11): Q is stronger than Q' just in case Q has greater information theoretic entropy (Shannon 1948) than Q'.

(11) $Q \subset Q' :\Leftrightarrow \mathrm{Ent}_{\mathrm{Pr}_s}(Q) > \mathrm{Ent}_{\mathrm{Pr}_s}(Q')$
where $\mathrm{Ent}_{\mathrm{Pr}}(Q) = \sum_{q \in Q} \mathrm{Pr}(q) \times \log_2(\frac{1}{\mathrm{Pr}(q)})$

The subscripts that Ent carries in (11) indicate that the entropy of a set of a proposition Q is defined relative to a probability mass function Pr with domain Q, that is, a function that maps each member of Q to a probability such that the probabilities in the range of Pr sum up to 1. The subscript s in Pr_s indicates that the ordering of questions is intended to be relative to

$Pr(p\cap q)$	$Pr(\neg p\cap q)$	$Pr(p\cap \neg q)$	$Pr(\neg p\cap \neg q)$	$Ent_{Pr}(Q1)$
0	0	0	1	0
$\frac{1}{16}$	$\frac{3}{16}$	$\frac{3}{16}$	$\frac{9}{16}$	1.62
$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	2
$\frac{9}{16}$	$\frac{3}{16}$	$\frac{3}{16}$	$\frac{1}{16}$	1.62
1	0	0	0	0

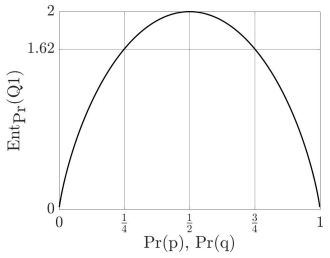


Figure 1: Question entropy as a function of the probabilities of exhaustive answers (table) and Hamblin answers (graph)

a probability mass function determined by the speaker's information state s. The entropy of a question partition Q relative to any given probability mass function Pr is a measure of how evenly Pr distributes the probability mass over the propositions in Q. Accordingly, the entropy of Q relative to Pr_s is a measure of the speaker's uncertainty about which member of Q is true.

To illustrate, the table in Figure 1 specifies the entropy relative to a few selected probabilities mass functions for the question partition Q1 defined in (12) below. The entropy is minimal (at 0) if all the probability mass is in one of the cells of the partition (first and fifth row) and maximal (here 2) if all the cells of the partition have equal probability (third row). For a probability mass function determined by the speaker's information state, the former case amounts to the speaker's beliefs entailing a complete answer to the question, while the latter case amounts to the speaker being maximally uncertainty or unbiased as to the question's true answer. Probability mass functions that are between those two extremes determine entropy values between 0 and the maximal entropy value (here 1.62, second and fourth row).

(12) $Q1 = \{p\cap q, \neg p\cap q, p\cap \neg q, \neg p\cap \neg q\}$

We are now interested in how the entropy of this question denotation depends on the probabilities of the Hamblin answers p and q. Assuming that p and q are independent and have equal probability, this dependency is as shown by the graph on the right-hand side of Figure 1. It can be read off this graph that the NPI condition is satisfiable for (3). To see why, suppose again that the domain is comprised of just a and b, and hence that the question semantics is as in (10) above. This semantics guarantees the truth of (13a), that is, that any two alternative Hamblin answers **S'**(a) and **S'**(b) are strictly stronger than the actual Hamblin answers **S**(a) and **S**(b), respectively. Probability theory in turn guarantees the truth of (13b), that is, that the probabilities of any two alternative Hamblin answers **S'**(a) and **S'**(b) are no greater than those of the actual Hamblin answers **S**(a) and **S**(b), respectively.

(13) a. $\forall$ **S'**[**S'** $\in [\![\text{said anything}]\!]^{ALT} \rightarrow$ **S'**(a) $\subset$ **S**(a) & **S'**(b) $\subset$ **S**(b)]
b. $\forall$ **S'**[**S'** $\in [\![\text{said anything}]\!]^{ALT} \rightarrow Pr(\mathbf{S}(a)) \geq Pr(\mathbf{S'}(a))$ & $Pr(\mathbf{S}(b)) \geq Pr(\mathbf{S'}(b))$]

Suppose now that $Pr_s(\mathbf{S}(a))$ and $Pr_s(\mathbf{S}(b))$ are equal and are both no greater than $\frac{1}{2}$ but greater than 0. Given (13b), since the actual Hamblin answers **S**(a) and **S**(b) are independent, we can then read off the graph in Figure 1 that any alternative question will have no greater entropy than the actual question. So, if the probabilities of all alternative Hamblin answers **S'**(a) and **S'**(b) are different from the probabilities of **S**(a) and **S**(b), (9) above will be true,

ensuring that the NPI condition (4b) is met. Hence the NPI condition is satisfiable, correctly permitting the acceptability of questions like (3).

4 The exhaustivity-licensing generalization derived

The analysis of NPI licensing by questions reviewed above, call it the *strength-as-entropy analysis*, captures one half of Guerzoni and Sharvit's (2007) exhaustivity-licensing generalization. Under the assumption that *wonder* and *know* select for the (strongly) exhaustive question meaning, the acceptable embedding cases in (1) and (2a) can satisfy the NPI condition (4b) in much the same way as the matrix question (3). By choosing the embedded question as the relevant syntactic domain, the demonstration of satisfiability for the matrix question (3) carries over. The only possible adjustment concerns the nature of the probability function relative to which entropy is calculated. In section 3, entropy was taken to be relative to a probability function determined by the speaker's information state s, while (1) and (2a) might instead make reference to attitude holder's information state.

But what about the other half of the exhaustivity-licensing generalization? What accounts for the unacceptability of (2b)? In brief commentary, Guerzoni and Sharvit (2007, 370) suggest that a strength-as-entropy analysis does not shed any light on the exhaustivity-licensing generalization. However, that assessment can be questioned. As noted in section 1, Heim (1994) presented observations indicating that the meaning of *surprise*, unlike the meaning of *know* (and presumably *wonder*), does not make reference to negated Hamblin answers. Under the present setup, this suggests that, instead of the (strongly) exhaustive semantics in (10), the embedded question in (2b) has the non-exhaustive semantics in (14), where negated Hamblin answers do not contribute to the membership of the answer set (Hamblin 1973, Karttunen 1977).

(14) a. $[\![\text{who said anything?}]\!] =$
$\{\mathbf{S}(\text{b}), \mathbf{S}(\text{a}), \mathbf{S}(\text{a})\cap\mathbf{S}(\text{b})\}$

b. $[\![\text{who said anything?}]\!]^{\text{ALT}} =$
$\{\ \{\mathbf{S'}(\text{b}), \mathbf{S'}(\text{a}), \mathbf{S'}(\text{a})\cap\mathbf{S'}(\text{b})\}\text{: } \mathbf{S'} \in [\![\text{said anything}]\!]^{\text{ALT}}\}$

How does a strength-as-entropy analysis apply under such a non-exhaustive question semantics? The first question is whether the notion of entropy is well-defined for the relevant meanings. The answer is that it can be. There exist probability mass functions that have non-exhaustive question meanings as their domain relative to which entropy can be calculated. For example, assuming that $\mathbf{S}(\text{a})$ and $\mathbf{S}(\text{b})$ are independent, so that $\Pr(\mathbf{S}(\text{a})\cap\mathbf{S}(\text{b})) = \Pr(\mathbf{S}(\text{a}))\times\Pr(\mathbf{S}(\text{b}))$, Pr is a probability mass function with domain (14a) if $\Pr(\mathbf{S}(\text{a})) = \frac{1}{2}$ and $\Pr(\mathbf{S}(\text{b})) = \frac{1}{3}$, since $\frac{1}{2} + \frac{1}{3} + \frac{1}{2} \times \frac{1}{3} = 1$. Hence (14a) has a well-defined entropy relative to this function (viz. 1.46).

Even so, however, under the strength-as-entropy analysis, such a non-exhaustive question semantics makes it impossible for the NPI condition to be met at the question level. That is, the strength-as-entropy analysis derives the exhaustivity-licensing condition introduced in section 1. To see why, note that the consequence of the NPI semantics stated in (13a) above entails (15a); therefore, in addition to (13b), probability theory guarantees (15b).

(15) a. $\forall\ \mathbf{S'}[\mathbf{S'} \in [\![\text{said anything}]\!]^{\text{ALT}} \rightarrow \mathbf{S'}(\text{a})\cap\mathbf{S'}(\text{b}) \subseteq \mathbf{S}(\text{a})\cap\mathbf{S}(\text{b})]$

b. $\forall\ \mathbf{S'}[\mathbf{S'} \in [\![\text{said anything}]\!]^{\text{ALT}} \rightarrow \Pr(\mathbf{S}(\text{a})\cap\mathbf{S}(\text{b})) \geq \Pr(\mathbf{S'}(\text{a})\cap\mathbf{S'}(\text{b}))\ ]$

So each member of the actual non-exhaustive question meaning (14a) is at least as likely as its counterpart in any of the alternatives. This has the following consequence. Suppose that

the probabilities of the members of the actual question meaning sum to 1. For the probabilities of members of an alternative question to sum to 1 as well, each of those members must have the very same probability as its actual counterpart. That is, in virtue of entailing (15), (13) also entails (16).

(16) $\Pr(\mathbf{S}(a)) + \Pr(\mathbf{S}(b)) + \Pr(\mathbf{S}(a) \cap \mathbf{S}(b)) = 1 \rightarrow \forall \mathbf{S'}[\mathbf{S'} \in [\![\text{said anything}]\!]^{ALT} \ \& \ \Pr(\mathbf{S'}(a)) + \Pr(\mathbf{S'}(b)) + \Pr(\mathbf{S'}(a) \cap \mathbf{S'}(b)) = 1 \rightarrow [\Pr(\mathbf{S'}(a)) = \Pr(\mathbf{S}(a)) \ \& \ \Pr(\mathbf{S'}(b)) = \Pr(\mathbf{S}(b)) \ \& \ \Pr(\mathbf{S'}(a) \cap \mathbf{S'}(b)) = \Pr(\mathbf{S}(a) \cap \mathbf{S}(b)) \ ]]$

Once again, the entropy of a question meaning is only defined relative to a probability mass function that has the question meaning as its domain, and this requires that the probabilities assigned to the members of a question meaning sum to 1. Therefore, since the entropy of a question meaning is determined by the probabilities assigned to the answers it contains, in the case at hand the entropy of any of the alternative question meaning, if defined, must equal the entropy of the actual question meaning. That is, (16) entails (17).

(17) $\text{Ent}_{\Pr}([\![\text{who said anything?}]\!])$ is defined $\rightarrow$ $[\forall Q'[Q' \in [\![\text{who said anything?}]\!]^{ALT} \ \& \ \text{Ent}_{\Pr}(Q') \text{ is defined} \rightarrow \text{Ent}_{\Pr}(Q') = \text{Ent}_{\Pr}([\![\text{who said anything?}]\!]) \ ]]$

An obvious consequence of (17), stated in (18), is that the actual question meaning in (14b) cannot be stronger than any of the alternative question meanings in (14b). Given that the set of alternative question meanings in (14b) is non-empty, (18) contradicts the requirement (9).

(18) $\forall Q' \ [Q' \in [\![\text{who said anything?}]\!]^{ALT} \rightarrow [\![\text{who said anything?}]\!] \not\subset Q' \ ]$

The non-exhaustive question meaning, in other words, cannot possibly reverse strength. This completes the argument that it is impossible for the non-exhaustive question meaning in (14) to satisfy the NPI condition (4b). The exhaustivity-licensing condition on question meanings, and hence Guerzoni and Sharvit's (2007) exhaustivity-licensing generalization, has been derived.

The strength-as-entropy analysis, then, is a possible account of the exhaustivity-licensing generalization. Is it the right account? In support of the analysis, the next section demonstrates that it makes additional welcome predictions about NPI licensing by question.

5 Support from presuppositional questions

Guerzoni and Sharvit (2014) report that in contrast to wh-questions, disjunctive questions (also known as alternative questions) never license NPIs. This generalization is illustrated by the unacceptability of (19).

(19) #Did Al say anything or Ben?

Schwarz (in press) observes that this is expected under the strength-as-entropy analysis, given independent observations about the interpretation of disjunctive questions. Such questions have been observed to carry a presupposition of existence and uniqueness (e.g., Karttunen and Peters 1976). For example, *Did Al talk or Ben*? presupposes that exactly one of Al and Ben talked. Under a partition semantics, this presupposition can be captured in the question meaning by expunging from it those answers that are inconsistent with the presupposition. This leads to the question meaning for (19) shown in (20). (20) excludes the answer that neither of Al and Ben said something as well as the answer that both did, encoding the existence and uniqueness presupposition as the disjunction of the remaining two propositions.

(20) a. $[\![$ did Al say anything or Ben? $]\!]$ =
$\{\text{–}\mathbf{S}(a)\cap\mathbf{S}(b), \mathbf{S}(a)\cap\text{–}\mathbf{S}(b)\}$

b. $[\![$ did Al say anything or Ben? $]\!]^{ALT}$ =
$\{ \{\text{–}\mathbf{S'}(a)\cap\mathbf{S'}(b), \mathbf{S'}(a)\cap\text{–}\mathbf{S'}(b)\}: \mathbf{S'} \in [\![$ say anything $]\!]^{ALT}\}$

Schwarz shows that, if the probabilities of the members of the actual question in (20a) sum to 1 and the same holds for the members of any alternative question in (20b), then the probabilities of the members of the alternative question must equal the probabilities of the corresponding propositions in the actual question. That is, (21) holds for (20).

(21) $Pr(\mathbf{S}(a)\cap\text{–}\mathbf{S}(b)) + Pr(\text{–}\mathbf{S}(a)\cap\mathbf{S}(b)) = 1 \rightarrow \forall\mathbf{S'}[\mathbf{S'} \in [\![$ said anything $]\!]^{ALT}$ &
$Pr(\mathbf{S'}(a)\cap\text{–}\mathbf{S'}(b)) + Pr(\text{–}\mathbf{S'}(a)\cap\mathbf{S'}(b)) = 1 \rightarrow [Pr(\mathbf{S}(a)\cap\text{–}\mathbf{S}(b)) = Pr(\mathbf{S'}(a)\cap\text{–}\mathbf{S'}(b))$ &
$Pr(\text{–}\mathbf{S}(a)\cap\mathbf{S}(b)) = Pr(\text{–}\mathbf{S'}(a)\cap\mathbf{S'}(b))$]]

As recorded in (22), this entails that the entropy of any of the alternative question meanings, if defined, must equal the entropy of the actual question meaning; (22) entails (23), that is, it entails that the actual question meaning in (20a) cannot be stronger than any of the alternative question meanings in (20b); and given that the set of alternative question meanings is non-empty, (23) contradicts (24). Under the present assumptions, then, the disjunctive question meaning cannot possibly reverse strength, hence cannot possibly satisfy the NPI condition (4b). So it is correctly predicted that such questions cannot license NPIs.

(22) $Ent_{Pr}([\![$ did Al say anything or Ben? $]\!])$ is defined $\rightarrow$
$[\forall Q'[Q' \in [\![$ did Al say anything or Ben? $]\!]^{ALT}$ & $Ent_{Pr}(Q')$ is defined $\rightarrow Ent_{Pr}(Q') =$
$Ent_{Pr}([\![$ did Al say anything or Ben? $]\!])$]

(23) $\forall Q'$ [$Q' \in [\![$ did Al say anything or Ben? $]\!]^{ALT} \rightarrow [\![$ did Al say anything or Ben? $]\!] \not\subset Q']$

(24) $\forall Q'$ [$Q' \in [\![$ did Al say anything or Ben? $]\!]^{ALT} \rightarrow [\![$ did Al say anything or Ben? $]\!] \subset Q']$

The statements in (21)–(23) are transparently parallel to those in (16)–(18) above. Under the strength-as-entropy analysis, then, questions with NPIs embedded under *surprise* and disjunctive questions with NPIs form a natural class. In both cases, the independently supported question meaning turns out to not allow any variation in entropy between the actual question and its alternatives generated by the NPI semantics, thereby rendering the satisfaction of the NPI condition impossible.

Are there other types of questions that that belong to this family? Singular *which*-questions (not discussed in Schwarz in press) are a natural candidate. Such questions, too, have been observed to carry a presupposition of existence and uniqueness (e.g., Higginbotham 1993, Dayal 1996). For example, *Which student talked*? is judged to presupposes that exactly one student talked. If the students are a and b, (25) should then have the very same semantics as (19), that is, the meaning given in (26).

(25) %Which student said anything?

(26) a. $[\![$ which student said anything? $]\!]$ =
$\{\mathbf{S}(a)\cap\text{–}\mathbf{S}(b), \text{–}\mathbf{S}(a)\cap\mathbf{S}(b)\}$

b. $[\![$ which student said anything? $]\!]^{ALT}$ =
$\{ \{\mathbf{S'}(a)\cap\text{–}\mathbf{S'}(b), \text{–}\mathbf{S'}(a)\cap\mathbf{S'}(b)\}: \mathbf{S'} \in [\![$ said anything $]\!]^{ALT}\}$

Under this semantics, for the very same reasons as (19), (25) necessarily violates the NPI

condition. Since this holds true more generally for any non-empty set of students (as the reader is invited to confirm), (25) is predicted to pattern with (19) in being judged unacceptable.

This prediction matches the judgments of some speakers. However as the diactritic % is meant to signal, others find such questions quite acceptable. In fact, without specifically discussing singular *which*-questions, Krifka (1995, 251) presents (27) as an example of NPI licensing by (information–seeking) wh-questions. Yet some speakers, those who reject (25), also reject (27).

(27) %Which student has ever been to China?

How might the present line of analysis accommodate this speaker variation? How could it be reconciled with speaker judgments that allow for NPIs to be licensed by singular *which*-questions? As a tentative answer, based on informants' reports about (25) and (27), I propose that singular *which*-questions do not in fact invariably carry a presupposition of existence and uniqueness. Specifically, I propose that the existence presupposition can be suspended. Hence I hypothesize that (25) and (27) are judged acceptable by speakers who interpret those questions as consistent with no student having talked and gone to China, respectively.

For those speakers, the question semantics in (26) is to be replaced with (28), where the proposition that neither a nor b said something is admitted as a member of the question set. As a consequence, this meaning now encodes a mere presupposition of uniqueness, rather than a presupposition of existence and uniqueness.

(28) a. $[\![\text{which student said anything?}]\!] =$
$\{\mathbf{S}(a)\cap\text{–}\mathbf{S}(b),\ \text{–}\mathbf{S}(a)\cap\mathbf{S}(b),\ \text{–}\mathbf{S}(a)\cap\text{–}\mathbf{S}(b)\}$

b. $[\![\text{which student said anything?}]\!]^{\text{ALT}} =$
$\{\ \{\mathbf{S'}(a)\cap\text{–}\mathbf{S'}(b),\ \text{–}\mathbf{S'}(a)\cap\mathbf{S'}(b),\ \text{–}\mathbf{S'}(a)\cap\text{–}\mathbf{S'}(b)\}\text{: } \mathbf{S'} \in [\![\text{said anything}]\!]^{\text{ALT}}\}$

Note that this revision suffices to render the NPI condition satisfiable in the case at hand. By reasoning familiar from section 3, this can be read off the graph in Figure 2, which plots the entropy of the question meaning Q2 in (29) below as a function of the probabilities of the Hamblin answers p and q, assuming that p and q are independent and have equal probability.[1]

(29) $Q2 = \{\text{–}p\cap q,\ p\cap\text{–}q,\ \text{–}p\cap\text{–}q\}$

In sum, provided the proposed interpretation of the singular *which*-question data is correct, the evidence from disjunctive questions and singular *which*-question strengthens the case for the strength-as-entropy analysis of the exhaustivity-licensing generalization, by demonstrating that variation of entropy values between the actual question and its alternatives is a necessary condition for NPI licensing.

6 Conclusion

The main result of this paper is that van Rooy's (2003) strength-as-entropy analysis of NPI licensing by questions can derive Guerzoni and Sharvit's (2007) exhaustivity-licensing generalization. Data from disjunctive questions and singular *which*-questions are proposed to provide further independent support for the approach.

[1] Note that the graph in Figure 2 only shows entropy values for Hamblin answer probabilities in the interval $[0, \frac{1}{2}]$. The reason for this (left for the reader to verify) is that the probabilities of the propositions in Q2 will not sum to 1 (hence the question entropy will not be defined) if the probabilities of the Hamblin answers p and q are greater than $\frac{1}{2}$.

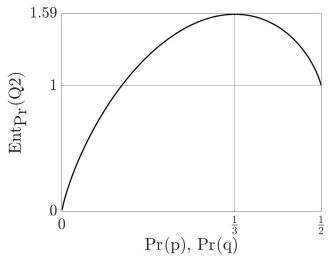


Figure 2: Entropy of a question that carries a uniqueness presupposition as a function of probabilities of Hamblin answers

A remaining question worth highlighting concerns the nature of the exhaustive/non-exhaustive distinction. The strength-as-entropy analysis explored here assumes that questions themselves are ambiguous (George 2011, Guerzoni and Sharvit 2014, Nicolae 2015, Theiler et al. 2016). However, this view has been challenged in recent work on question embedding, most notably Klinedinst and Rothschild (2011) and Uegaki (2015). It remains to be seen whether or how the arguments given in those works can be reconciled with the present proposal.

References

Chierchia, Gennaro. 2013. *Logic in Grammar: Polarity, Free Choice, and Intervention*. Oxford: Oxford University Press.

Dayal, Veneeta. 1996. *Locality in wh quantification*. Dordrecht: Kluwer.

George, Benjamin Ross. 2011. Question embedding and the semantics of answers. Doctoral Dissertation, University of California, Los Angeles.

Groenendijk, Joroen, and Martin Stokhof. 1982. Semantic analysis of wh-complements. *Linguistics and Philosophy* 5:175–233.

Guerzoni, Elena, and Yael Sharvit. 2007. A question of strength: on NPIs in interrogative clauses. *Linguistics and Philosophy* 30:361–391.

Guerzoni, Elena, and Yael Sharvit. 2014. *Whether or not anything* but not *whether anything or not*. In *The Art and Craft of Semantics: a Festschrift for Irene Heim*, ed. Luka Crnič and Uli Sauerland, 199–224. Cambride, MA: MIT Working Papers in Linguistics (MITWPL).

Hamblin, Charles L. 1973. Questions in Montague English. *Foundations of Language* 10:41–53.

Heim, Irene. 1994. Interrogative semantics and karttunens semantics for know. In *Proceedings of the Ninth Annual Conference of the Israeli Association for Theoretical Linguistics and the Workshop on Discourse*, ed. Rhona Buchalla and Anita Mitwoch, volume 1, 128–144.

Higginbotham, James. 1993. Interrogatives. In *The View from Building 20*, ed. Kenneth Hale and Jay Keyser. Cambride, MA: MIT Press.

Kadmon, Nirit, and Fred Landman. 1990. Polarity sensitive *any* and free choice *any*. In *Proceedings of the 7th Amsterdam Colloquium*, ed. Martin Stokhof and L. Toorenvliet, 227–251. Institute for Language Logic and Information.

Kadmon, Nirit, and Fred Landman. 1993. Any. *Linguistics and Philosophy* 16:353–422.

Karttunen, Lauri. 1977. Syntax and semantics of questions. *Linguistics and Philosophy* 1:3–44.

Karttunen, Lauri, and Stanley Peters. 1976. What indirect questions conventionally implicate. In *Papers from the 12th Regional Meeting of the Chicago Linguistic Society Chicago*, ed. Salikoko S. Mufwene, 351–368. Chicago, Illinois.

Klima, Edward S. 1964. Negation in English. In *The Structure of Language*, ed. Jerry A. Fodor and Jerrold J. Katz, 246–323. New Jersey: Prentice-Hall.

Klinedinst, Nathan, and Daniel Rothschild. 2011. Exhaustivity in questions with non-factives. *Semantics and Pragmatics* 4:1–23.

Krifka, Manfred. 1995. The semantics and pragmatics of polarity items. *Linguistic Analysis* 25:209–257.

Ladusaw, William. 1979. Polarity sensitivity as inherent scope relations. Doctoral Dissertation, University of Texas at Austin.

Lahiri, Utpal. 1998. Focus and negative polarity in Hindi. *Natural Language Semantics* 6:57–123.

Nicolae, Andreea C. 2015. Questions with NPIs. *Natural Language Semantics* 23:21–76.

Rooth, Mats. 1985. Association with focus. Doctoral Dissertation, University of Massachusetts Amherst.

van Rooy, Robert. 2003. Negative polarity items in questions: Strength as relevance. *Journal of Semantics* 20:239–273.

Schwarz, Bernhard. in press. Negative polarity items: a case for questions as licensers. In *Proceedings of Semantics and Linguistic Theory (SALT) 27*, ed. Dan Burgdorf and Jacob Collard, 230–247. Ithaca, NY: Cornell University.

Shannon, Claude Elwood. 1948. A mathematical theory of communication. *Bell System Technical Journal* 27:379–423.

Theiler, Nadine, Floris Roelofsen, and Maria Aloni. 2016. Truthful resolutions: A new perspective on false-answer sensitivity. In *Proceedings of Semantics and Linguistic Theory (SALT) 26*, ed. Mary Moroney, Carol-Rose Little, Jacob Collard, and Dan Burgdorf, 122–141. CLC Publications.

Uegaki, Wataru. 2015. Interpreting questions under attitudes. Doctoral Dissertation, Massachusetts Institute of Technology, Cambridge, Massachusetts. URL https://dspace.mit.edu/handle/1721.1/99318.

A Stalnakerian Analysis of Metafictive Statements

Merel Semeijn

University of Groningen, Groningen, The Netherlands

Abstract

Because Stalnaker's common ground framework is focussed on cooperative information exchange, it is challenging to model fictional discourse. To this end, I develop an extension of Stalnaker's analysis of assertion that adds a temporary workspace to the common ground. I argue that my framework models metafictive discourse better than competing approaches that are based on adding unofficial common grounds.

1 Fiction in a Stalnakerian Framework

In Stalnaker's (1970) widely adopted pragmatic framework, assertions are modelled as updates of the 'common ground' (i.e. the set of propositions that are mutually presupposed by speaker and addressee in a conversation). When I assert "It is raining", I propose to update the common ground between me and my addressee with the proposition that it is raining. Because Stalnaker focussed on cooperative information exchanges (i.e. conversations in which gradually more and more information is added to the common ground because more and more propositions are asserted and accepted), it is challenging to model speech acts that do not seem to follow standard Gricean maxims, such as telling a lie or a fictional story which involves asserting things that you know not to be (literally) true.[1]

In this paper I propose a new approach to modelling fictional discourse in a Stalnakerian framework. I first discuss and present a formalization of the 'unofficial common ground accounts' offered by Stokke (2013) and Eckardt (2014), in section 2. Then, in section 3, I argue that Stokke and Eckardt run into difficulties modelling so-called 'metafictive statements' (i.e. reports on the content of a fictional work). In section 4, I introduce my own Stalnakerian analysis of fictional statements and argue that it adequately models metafictive statements. Instead of unofficial common grounds, my account involves what I call a temporary 'workspace'. Hence I dub my own account the 'workspace account'. Lastly, in section 5, I present a remaining challenge for my account and suggest directions for further research.

2 Unofficial Common Ground Accounts

A sharp contrast is drawn between assertions on the one hand and fictional statements on the other, in both Eckardt's linguistically motivated approach and Stokke's philosophically motivated approach to modelling fiction in a Stalnakerian framework. Assertions are updates of the 'official common ground': the set of mutually presupposed propositions concerning actual states of affairs. Fictional statements however, are proposals to update or create an 'unofficial common ground' related to the relevant fictional work: the set of propositions mutually presupposed by the addressee and author[2] of a story. As we engage in a fictional narrative the relevant unofficial

[1] Although these speech act are traditionally treated on a par, there are crucial differences (cf. Maier, forthcoming) such as that stereotypical lies, contrary to fictional statements, involve an intention to deceive.

[2] Hence, it is unofficial common ground between me and Tolkien that wizards exist. Alternatively, we can analyse this as part of the *official* common ground between the narrator and narratee encoded in the fiction. I will not further develop the role of the narrator in this paper.

common ground is updated with propositions expressing the content of the fictional story. Take fictional statement (1), taken from Tolkien's novel *The Lord of the Rings*:

(1) Frodo had a very trying time that afternoon.

As I read (1), I update the unofficial common ground specifically related to Tolkien's narrative with the proposition that Frodo had a very trying time on a particular afternoon.

Because we normally engage in different fictional narratives, a typical 'complete common ground' will contain one official common ground concerning actual states of affairs, and several unofficial common grounds related to different fictions (e.g. a *The Lord of the Rings* common ground, a *Harry Potter* common ground, a *Pride and Prejudice* common ground, etc.).[3] The complete common ground (C) can thus be represented as a n-tuple of one official common ground (C_0), and several numbered unofficial common grounds ($C_1, ..., C_n$):

$$C = \langle C_0, C_1, ..., C_n \rangle$$

Assertions are defined as proposals to update ($*$) the official common ground:

$$C +_A p = \langle C_0 * p, C_1, , C_n \rangle$$

Importantly, updating a common ground with some proposition p is usually formalized as $C \cup \{p\}$ rather than $C * p$, when common grounds are defined as sets of propositions ($C \cap p$ when common grounds are defined as sets of possible worlds). However, especially with fictional statements, simply unionizing sets of propositions leads to inconsistent common grounds. Further research is needed to determine what operator is suitable for such inconsistent updates (e.g. a belief revision-, Lewisian- or probability distribution-operator). $*$ may denote any such operator.

Fictional statements are defined as proposals to update an unofficial common ground. Here we must distinguish between two cases: Either a fictional statement is a proposal to update an already existing unofficial common ground (e.g. when continuing to read *The Lord of the Rings*), or a fictional statement is a proposal to create a new unofficial common ground (C_{BASE}) and update this common ground (e.g. when starting to read a new fictional novel):

$$C +_{F^i} p = \begin{cases} \langle C_0, C_1, ..., C_{i-1}, C_i * p, C_{i+1}, ..., C_n \rangle \text{if } 1 \leq i \leq n, \\ \langle C_0, C_1, ..., C_n, C_{BASE} * p \rangle, \text{ otherwise.} \end{cases}$$

This formalization raises the question what exactly is the content of C_{BASE}. In other words, what is mutually presupposed by the addressee and author of a fictional story when starting to engage in a new fictional narrative? Is C_{BASE} a copy of the official common ground ($C_{BASE} = C_0$), a tabula rasa ($C_{BASE} = \varnothing$), or something in between? Answering this question is outside the scope of this paper (but see for instance Lewis (1978), Ryan (1980) or Lamarque (1990)). In the formalizations presented in this paper I assume that C_{BASE} is a copy of the official common ground between addressee and author and hence contains all mutually presupposed propositions concerning actual states of affairs. Assuming C_{BASE} is a tabula rasa, or something in between a tabula rasa and a copy of the official common ground, is also compatible with my model but would lead to different formalizations.

[3] Alternatively, following the 'fragmented mind' programme (David), one could formulate an account involving *one* compartmentalized common ground. What are unofficial common grounds in Stokke's and Eckardt's accounts, are different compartments related to different fictions in this framework. Beliefs concerning actual states of affairs (part of the official common ground in the unofficial common ground accounts) are also structured in compartments of the same common ground.

3 Metafictive Discourse

So far I have focussed on Stokke's and Eckardt's analysis of fictional statements. However, there are also other types of discourse connected to our engagement with fiction. In this section, I argue that both Stokke and Eckardt run into difficulties concerning temporality and ascribing intuitively correct truth-values when modelling 'metafictive discourse'.

3.1 Conflicting Intuitions

Philosophers of fiction draw a distinction between (in Currie's (1990) terminology) 'fictional statements' which are taken directly from some fictional work (e.g. (1)) and 'metafictive statements'[4] which are statements that *are* about the content of a fictional work but that are not taken from it. For instance, after reading *The Lord of the Rings* people may start the following discussion:

A: Did you know that Frodo was adopted by his uncle?
B: Actually, Bilbo is Frodo's cousin.

So, after reading *The Lord of the Rings*, although I no longer accept or imagine the content that I entertained while engaging with the fictional statements of the narrative, I do not forget it. I remember that Bilbo is Frodo's cousin and I thus make a metafictive statement (2) about the content of a fictional work after engaging with it:

(2) Bilbo is Frodo's cousin.[5]

It seems that there are two conflicting intuitions that we want to account for when modelling our engagement with fiction: First, the acceptance of fictional truths is temporary (only for the purpose and duration of the conversation) and second, we do retain information about the fictional content *after* engaging with the narrative. I for example only momentarily accept propositions such as that wizards and hobbits exist for the purpose of reading *The Lord of the Rings*. However, even after engaging in the narrative, I do remember that Bilbo is Frodo's cousin and would correct someone who stated otherwise. Stokke and Eckardt run into difficulties trying to account for these conflicting intuitions.

Here it is important to point out a crucial difference between Stokke's and Eckardt's theories. In Eckardt's framework unofficial common grounds are non-temporal; once they come into existence they continue to exist. Because we engage with many different fictions over the course of our lives, a typical complete common ground consists of one official common ground and an ever-growing number of coexisting unofficial common grounds that are continuously accessible. In this framework we can treat metafictive statements like (2) as fictional statements: (2) is a proposal to update the unofficial common ground related to *The Lord of the Rings* with the proposition that Bilbo is Frodo's cousin. However, since unofficial common grounds are non-temporal in Eckardt's framework (i.e. they continue to be accessible after engaging with the

[4] There is the threat of a terminological confusion here: metafictive statements differ from so-called 'metafictional statements' (e.g. "Frodo is a fictional character" or "Frodo does not exist"). Further research will have to determine how to model metafictional statements in a Stalnakerian framework.

[5] We could also imagine this sentence occurring in *The Lord of the Rings*. This shows that whether a utterance is a fictional statement or a metafictive statement is largely a matter of context; the same sentence can function as a fictional statement (when found in a fictional work) or as a metafictive statement (when found in a discussion on the content of the fictional work such as (2)).

fictional narrative), Eckardt cannot account for the first intuition that fictional truths are only accepted temporarily.

Alternatively, Stokke analyses unofficial common grounds as essentially temporal, contrasting them with "more permanent, 'official', common grounds" (Stokke, 2013, p. 53), to account for the intuition that fictional statements are only accepted temporarily. He explains the use of metafictive discourse[6] by claiming that "an unofficial common ground need not be temporary in the sense of lasting a short time. There are arguably [unofficial] common grounds [that] continue to be operative for a very long time" (Stokke, 2013, p. 55). In other words, like in Eckardt's framework, metafictive statements are treated on a par with fictional statements (i.e. they are proposals to update an unofficial common ground). Therefore, in order to account for metafictive discourse, Stokke has to admit that unofficial common grounds remain operative long after engaging with the fictional narrative. But in what sense are unofficial common grounds that remain accessible after engaging with a fictional narrative (as in Eckardt's theory) still temporal? It seems that Stokke runs into difficulties trying to account for both intuitions described above, ending up with unofficial common grounds that are both essentially temporal and continuously operative.

3.2 Lewis' fiction-operator

Now that the issues concerning temporality raised by metafictive statements for the unofficial common ground accounts have been discussed, let's further examine the nature of metafictive statements. I argue that the analysis of metafictive statements that Stokke and Eckardt adhere to runs into difficulties with ascribing intuitively correct truth-values.

A prominent theory in the philosophy of fiction is that a fictional statement p (e.g. (1)) is a truth-valueless mandate or proposal from the author to imagine p.[7] The nature of *metafictive* statements (e.g. (2)) on the other hand is still a matter of debate. Currie (1990) and others (e.g. Zucchi (2017) and Ninan (2017)) claim that metafictive statements are not mandates to imagine but rather abbreviations of assertions about the content of particular fictional works under a Lewisian (1978) 'In (the worlds compatible with) fiction x'-operator. Hence, (2) is actually an abbreviation of metafictive statement (3):

(3) In *the Lord of the Rings*, Bilbo is Frodo's cousin.

In other words, metafictive statements (with overt or covert fiction-operators) are meant to make assertions about actual states of affairs in the world (i.e. the content of a particular work of fiction) and are true or false depending on how the actual world is.

By contrast, Recanati (2002), Evans (1982) and Walton (1990) claim that both metafictive statements such as (2) *and* metafictive statements with an overt fiction-operator such as (3) are, like fictional statements, mandates to imagine the described events. They are truth-valueless invitations to continue the pretense of the *The Lord of the Rings* stories. The unofficial common ground framework, by treating fictional statements and metafictive statements on a par, seems to assume this analysis of metafictive statements.[8] There are independent reasons to prefer

[6] Stokke discusses the statement "Hobbits have hairy feet." (Stokke, 2013, p. 55) as an answer to the question "Who has hairy feet?". Stokke's example is somewhat complicated; it is not immediately clear whether the statement occurs in the context of a discussion on the content of *The Lord of the Rings*. However, we can interpret it as a metafictive statement (i.e. a statement about the content of *The Lord of the Rings*) because it is perfectly reasonable to reply: "Okay, true... But let's not talk about *The Lord of the Rings* right now."

[7] This consensus view has recently been challenged by Matravers (2014). See section 4.1.

[8] Stokke only discusses metafictive statements without overt fiction-operator. Possibly, metafictive statements with overt fiction-operator can be analysed as operating on the official common ground.

Currie's analysis, related to the behaviour of indexicals (See Zucchi (2017)). More importantly, the Recanati/Evans/Walton view does not allow us to ascribe truth values to metafictive statements; they are, like fictional statements, truth-valueless mandates to imagine. However, as Currie (1990) and Zucchi (2017) argue, we intuitively *do* want to maintain that "(In *The Lord of the Rings,*) Bilbo is Frodo's cousin." is a true statement. In this paper I follow Currie's analysis of metafictive discourse.

4 The Workspace Account

Now that I have discussed the unofficial common ground accounts and the challenge posed by metafictive discourse, I turn to my own version of a Stalnakerian model of fictional discourse. I will first present my Stalnakerian model of fictional statements and assertions (section 4.1). Second, I will discuss my analysis of metafictive statements and argue that it avoids the difficulties raised for the unofficial common ground accounts (section 4.2).

4.1 Fictional Statements and Assertions

As I mentioned in the introduction, my account involves a concept similar to the unofficial common ground. Following Stokke and Eckardt, I analyse fictional statements as proposals to update a common ground separate from the official common ground, that contains the shared presuppositions between author and addressee of the fictional story. However, in my account, this common ground truly temporal; it is operative or active solely for the purpose and solely for the duration of the fictional discourse. I hence dub it the 'workspace'. As soon as we stop entertaining the propositions of some fictional narrative (e.g. as soon as I stop reading *The Lord of the Rings*), the content of the workspace evaporates and it becomes inactive again.

Using a recent insight from the philosophy of fiction due to Matravers (2014) I claim that this is the case for both fictional and non-fictional discourse. Matravers challenges the widely adopted view that whereas nonfictional truths are to be believed, fictional truths are to be imagined. In fact, our primary engagement with fictional narratives (e.g. *The Lord of the Rings*) involves the essentially the same cognitive processes as our primary engagement with nonfictional narratives (e.g. a vivid biography); whether a narrative is fictional or non-fictional, entertaining its content involves the same cognitive mechanisms. Likewise, in my framework common ground updates are formalized as a two-step algorithm where the first step – updating a workspace – is uniform for fiction and non-fiction.[9] I define this first step as an update ($*$) on an ordered pair consisting of an official common ground (C) and a workspace (W) (which, when engaging in a new narrative, is inactive up to that point):

$$\langle C, W\rangle + p = \langle C, W * p\rangle$$

With the first update (i.e. with the first proposition of the narrative we are entertaining) the workspace becomes active. It then remains active during following updates caused by the same, possibly multi-sentence, discourse. In other words, when entertaining propositions from some narrative (e.g. *The Lord of the Rings* or a newspaper article), I activate the workspace with my first update and then continue to further update this workspace. When I stop entertaining propositions from this narrative (i.e. as I stop reading or listening), the workspace loses its content and becomes inactive again. As I subsequently engage in a new narrative (e.g. *Harry Potter*), I again update, and thereby activate, the same workspace.

[9] A similar idea is developed in Kamp's (2016) mentalistic framework. Kamp introduces a compartment (K_{dis}) for the neutral place where we build representations of the content of the current discourse before forming judgements about the truth of the propositions expressed by K_{dis}.

What differentiates assertions from fictional statements is how, at the end of the discourse, they update the official common ground: whether the content of the updated workspace is adopted as belief (for nonfiction) or as metafictive belief (for fiction). Hence, in the second step of the algorithm, I define assertions and fictional statements as different 'closure operations' that take an ordered pair $\langle C, W\rangle$ containing an updated, active workspace, and return an ordered pair with a new official common ground and an inactive workspace.

In the case of assertion the updated workspace is adopted as the new official common ground. As discussed in 2, I assume that an inactive workspace is a copy of the current official common ground (instead of for instance a tabula rasa) so that asserting a proposition p boils down to updating the official common ground C to $C * p$ (as in the traditional Stalnakerian framework):

$$\text{Assertive closure: } \langle C, W\rangle \rightarrow \langle W, W\rangle$$

Assuming we are keeping track of n fictions $(1, ..., n)$, fictional statements return an ordered pair in which the updated workspace is added to the original official common ground under an 'In fiction i'-operator. This is meant to model the fact that once a fictional discourse ends (e.g. once I put down *The Lord of the Rings*) participants are no longer invited to imagine any propositions but do maintain metafictive beliefs about the content of the fictional narrative (e.g. In *The Lord of the Rings*, Frodo is adopted by his cousin). In the formalization below the 'In fiction i'-operator ($\Box_i$) takes as its argument the proposition $\cap W$, which is the intersection of the propositions in W:

$$\text{Fictive}^i \text{ closure: } \langle C, W\rangle \rightarrow \langle C \cup \{\Box_i(\cap W)\}, C \cup \{\Box_i(\cap W)\}\rangle$$

So after engaging with some fictional narrative i, the official common ground will contain metafictive beliefs concerning i. The workspace becomes an inactive copy of this official common ground (which now also contains metafictive beliefs). Because we normally engage in different fictional narratives, a typical official common ground will contain (apart from beliefs about the actual world) metafictive beliefs about several distinct fictions under different 'In fiction i'-operators. In this sense there are in fact multiple different fictive closure operators related to different fictional works.

In sum, the workspace account analyses assertions as proposals to update a workspace and as a result of that adopt the workspace as the new official common ground (i.e. as belief). Fictional statements are proposals to update a workspace and as a result of that update the official common ground with the content of the workspace under the relevant 'In fiction i'-operator (i.e. with metafictive beliefs). Rather than using different update rules for fictional statements and assertions (as Stokke and Eckardt do), I thus propose a uniform workspace update, along with distinct assertive and fictive closure operations.

4.2 Metafictive Statements

Now that I have presented my version of a Stalnakerian model of fictional statements, in this section, I present my analysis of metafictive statements. I argue that my account avoids the difficulties described in section 3 (concerning temporality and ascribing intuitively correct truth-values) that Stokke and Eckardt run into.

I claim that as a result of entertaining fictional propositions, we add metafictive beliefs to the *official* common ground. Metafictive statements are reports on these metafictive beliefs. Hence, I model a metafictive statement such as (2) or (3) (i.e. with overt or covert 'In fiction i'-operator) as a proposal to update the *official* common ground with (3). Or, in other words, as a plain assertion about actual states of affairs. Any arbitrary metafictive proposition p consists

of an 'In fiction i'-operator related to some fiction i ($\Box_i$), and some proposition (q): $p = \Box_i q$. We can thus represent metafictive statements by substituting p for $\Box_i q$ in the two-step algorithm for assertions. First, we update the workspace with $\Box_i q$:

$$\langle C, W\rangle + \Box_i q = \langle C, W * \Box_i q\rangle$$

At the end of the (possibly multi-sentence) discourse about the content of fiction i, we perform regular assertive closure: $\langle C, W\rangle \rightarrow \langle W, W\rangle$ by which the updated workspace is adopted as the new official common ground (which now also contains $\Box_i q$). In other words, engaging in a fictional narrative i or engaging in a discussion on the content of i are two distinct ways of obtaining the same result: metafictive beliefs concerning i. The intuitive difference between the two processes lies in what kind of propositions you entertain (i.e. update your workspace with) during the discourse; whether you entertain propositions such as "Bilbo is Frodo's cousin" or propositions such as "In *The Lord of the Rings*, Bilbo is Frodo's cousin".

By analysing metafictive statements as plain assertions about actual states of affairs, I adopt Currie's view of metafictive statements. Hence I allow, unlike Stokke and Eckardt, that we ascribe truth-values to metafictive statements such as (3) just as we do with assertions. My account also adequately models the temporary acceptance of fictional propositions. After engaging in *The Lord of the Rings*, the fictional content of our workspace evaporates. Hence we accept propositions, such as that wizards exist, only temporarily. The metafictive beliefs that we subsequently adopt become part of the official common ground and are therefore more 'permanent' (i.e. as permanent as ordinary beliefs). Thus, after engaging in *The Lord of the Rings* I do remember that (in *The Lord of the Rings*,) Bilbo is Frodo's cousin. When engaging in metafictive discourse, we make regular assertions and therefore no 'permanent' unofficial common ground or workspace related to *The Lord of the Rings* is called for.

5 Picking up Where We Left of

In this section I discuss a challenge for my account: how do we model continuing to engage in a fictional narrative after taking a break? I suggest two directions in which to further develop the workspace account and hence meet this challenge.

A feature of my model is that after engaging in a fictional narrative, all that we are left with are metafictive beliefs. It is therefore unclear how we can for instance interpret "Gollum [...] held aloft the ring" after taking a break from reading *The Lord of the Rings*. What ring is Tolkien writing about? The new workspace will have to contain the propositions that were included in the original workspace just before fictive closure (e.g. a description of some unique ring) in order to account for such anaphoric links. However, the official common ground (and hence the new workspace) only contains metafictive propositions based on this original workspace.[10] My account is in need of some further mechanism to explain how we are able to retrieve the relevant propositions embedded under 'In fiction i'-operators.

A possible solution is to claim that, apart from adding metafictive propositions to the official common ground, fictive closure also involves retaining a copy of the updated workspace (which is adopted as new workspace when continuing in the same narrative). This results in a theory resembling Stokke's and Eckardt's accounts, involving (something akin to) unofficial common grounds. However, this move invites the problems concerning temporality associated with the

[10]This problem does not arise with non-fictional narratives because with assertive closure we adopt the updated workspace as the new official common ground. Hence, when continuing in a non-fictional narrative, the new workspace will contain (at least) all propositions that were included in the original workspace.

unofficial common ground accounts; if we maintain that readers of *The Lord of the Rings* hold onto a *The Lord of the Rings*-workspace containing propositions such as that wizards exist, we no longer account for the intuition that we accept such fictional propositions only temporarily.

A more promising solution is to maintain that after engaging in a fictional narrative we *only* retain metafictive beliefs. Hence a further mechanism is needed to explain how we can, when continuing in a familiar narrative, fill in our workspace appropriately, based on the available metafictive propositions in the official common ground. At first sight this seems like a straightforward task. In fictive closure you copy the updated workspace in its entirety and add it to the official common ground under an 'In fiction i'-operator. So, when you continue to engage in a fictional narrative after taking a break, you simply reverse the fictive closure (i.e. perform 'fictive opening'): identify the relevant fiction-operator and copy everything that is under this operator to the workspace.

The difficulty with this approach is that in Stalnaker's framework common grounds have no structure; they are simply sets of propositions (which in turn are sets of possible worlds) that determine a unique set of possible worlds. When we perform fictive closure we update the official common ground with (metafictive) propositions and thus further limit this set of possible worlds (e.g. we exclude worlds in which in the novel *The Lord of the Rings*, Bilbo is not Frodo's cousin). In other words, there is no 'metafictive *The Lord of the Rings*' marker in the official common ground and hence no straightforward mechanism to select the appropriate propositions to perform fictive opening.

A framework that does place structure on propositions and hence allows for metafictive markers, is the so-called 'structured propositions' framework (See for instance Soames (1985) and Cresswell (1985)). Propositions are not sets of possible worlds, but complex entities with a structure similar to the sentences that expresses them and with constituents that carry the semantic values of expressions occurring in these sentences. For example, in Soames' neo-Russellian approach the sentence "Scott does not run" expresses the following proposition:

$$< NEG, << s >, R >>$$

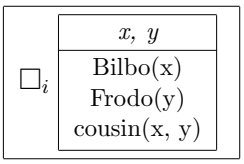
Here, s is Scott, R is the property of running and NEG is the truth function for negation. Thus, the negation operator is a distinct constituent of the proposition expressed. We can analyse the 'In fiction i'-operator in metafictive statements in a similar fashion. The following is a simplified representation of the proposition expressed by metafictive statement (3):

$$< \Box_i, << b, f >, C >>$$

Here, b is Bilbo, f is Frodo, C is the property of being someone's cousin and $\Box_i$ is the 'In *The Lord of the Rings*'-operator. In this way we place structure on propositions that allows for metafictive markers and hence enables us to perform fictive opening.

Perhaps a more suitable framework for dealing with the problem of anaphoric links (described above) is DRT (Discourse Representation Theory) developed by Kamp (1981)[11], because it is both structured and dynamic. In this framework NP's in a discourse are mapped to 'discourse referents' that are placed under several conditions in a DRS (Discourse Representation Structure). For example (3) is represented as follows:

[11] A similar theory has been developed by Heim (1982) independently.



Hereby we place structure on propositions that includes metafictive markers and hence allows us to to select the appropriate propositions to perform fictive opening.

6 Conclusions

I have argued that Stokke's and Eckardt's unofficial common ground accounts, in which fictional statements are proposals to update a separate unofficial common ground, run into difficulties with modelling metafictive discourse. I have presented an alternative Stalnakerian analysis of fictional statements: the workspace account. Rather than using different update rules for fictional statements and assertions, I propose a uniform workspace update along with distinct assertive and fictive closure operations. I have argued that my account avoids the difficulties with modelling metafictive discourse associated with the unofficial common ground accounts. A standing challenge is to adequately model how we can continue to engage in a fictional narrative after taking a break.

References

Cresswell, M. J. (1985). *Structured meanings: The Semantics of Propositional Attitudes*. MIT Press.

Currie, G. (1990). *The Nature of Fiction*. Cambridge University Press.

David, M. The Fragmented Mind: Agency, Rationality, Belief. http://fragmentationproject.uni-graz.at/.

Eckardt, R. (2014). *The Semantics of Free Indirect Discourse*, volume 31. Brill Publishers.

Evans, G. (1982). *The Varieties of Reference*, volume 10. Oxford University Press.

Heim, I. (1982). *The Semantics of Definite and Indefinite Noun Phrases*. PhD thesis, UMass Amherst.

Kamp, H. (1981). A theory of truth and semantic representation. In Groenendijk, J. A. G., Janssen, T. M. V., and Stokhof, M. B. J., editors, *Formal Methods in the Study of Language, Part 1*, pages 277–322. Blackwell Publishers Ltd, Oxford, UK.

Kamp, H. (2016). Entity representations and articulated contexts: An exploration of the semantics and pragmatics of definite noun phrases. Working paper.

Lamarque, P. (1990). Reasoning to what is true in fiction. *Argumentation*, 4(3):333–346.

Lewis, D. (1978). Truth in Fiction. *American Philosophical Quarterly*, 15(1):37–46.

Maier, E. Lying and fiction. In Meibauer, J., editor, *The Oxford Handbook of Lying*. Oxford: Oxford University Press, forthcoming.

Matravers, D. (2014). *Fiction and Narrative*. Oxford University Press.

Ninan, D. (2017). Names in fiction. *Theoretical Linguistics*, 43(1-2).

Recanati, F. (2002). Varieties of simulation. In Proust, J. and Dokic, J., editors, *Simulation and Knowledge of Action*, pages 151–171. John Benjamins Publishing.

Ryan, M. L. (1980). Fiction, non-factuals, and the principle of minimal departure. *Topics in Catalysis*, 9(4):403–422.

Soames, S. (1985). Lost innocence. *Linguistics and Philosophy*, 8(1):59–71.

Stalnaker, R. C. (1970). Pragmatics. *Synthese*, 22(1-2):272–289.

Stokke, A. (2013). Lying and asserting. *Journal of Philosophy*, 110(1):33–60.

Walton, K. L. (1990). *Mimesis as Make-Believe: On the Foundations of the Representational Arts*. Harvard University Press.

Zucchi, S. (2017). Games of make-believe and factual information. *Theoretical Linguistics*, 43(1-2).

Towards a semantic typology of specific determiners*

Alexandra Simonenko

Research Foundation Flanders & Ghent University, Ghent, Belgium
alexandra.simonenko@ugent.be

Abstract

This paper investigates properties of a class of determiners which can be loosely labelled specific in that their distribution falls in between maximally-quantifying definite determiners and indefinites which only contribute existential quantification. Based on a sample which includes data from Buryat, Komi, Mari, Mordvin, and Turkish, I propose that suffixal determiners form a semantically natural class in that their distribution can be modelled by means of a relational component in the semantics of the determiners which relates the denotation of the noun to an antecedent. I derive the observed distributional differences between languages from the range of values available for the interpretation of this component. In particular, whether a relation of identity falls within the range of values has consequences for whether a suffixal determiner triggers existence presupposition, which, in turn, has consequences both for the interpretation of the DP in question and for the inter-paradigm competition in a language.

1 Introduction

Determiners, understood as morphemes relating the denotation of a noun phrase to the more general truth- and felicity conditions, can be divided into two large classes on semantic grounds with respect to whether they trigger maximal quantification of the kind associated with Germanic definite determiners in the Sharvy-based developments of the Fregian/Russellian tradition.[1]

Maximally quantifying determiners, in a morphologically free-standing or affixal form, are found, for instance, in Romance, Germanic, Semitic, Albanian, and Greek. Leaving this class aside, this paper focuses on non-maximally quantifying determiners which relate the denotation of the noun to some contextually specified individual using an umbrella term of specificity, to be defined more precisely in the course of the discussion. Using a classic example from the seminal paper of Enç (1991) to set the stage, consider (1) where the use of the marker *-ı* implies that the denotation of *iki kız-ı* "two girls" is a subset of the group of individuals which verifies the truthfulness of the preceding sentence with the expression *birkaç çocuk*. In the absence of the marker, an inference arises that these are some unrelated girls.[2]

(1) Oda-m-a birkaç çocuk gir-di. Iki kız-ı tanı-yor-du-m.
room-1SG-DAT several child enter-PST.3SG two girl-ACC know-IMPF-PST-1SG
"Several children entered my room. I knew two (of the) girls." [Turkish], Enç (1991, 6)

*I am extremely thankful to Metin Bagriacik, Gyrylma Bato-Munhoevna Bazarova, Viktorija Viktorovna Batorova, Anastasia Artemovna Bormatova, Ojuna Bubeevna Budaeva, Emilia Philippovna Khozyinova, Zinaida Vetkeevna Klucheva, Sepideh Mortazavinia, Galina Gennadjevna Pushkina, Irina Valerjevna Shabalina, and Anastasia Ivanovna Vaslyaeva for language consultancy. I thank the anonymous reviewers of the Amsterdam Colloquium, the audience of the Dialing seminar at UGhent for their comments, and Svetlana Toldova for stimulating discussions.

[1] As in most cases of quantification in natural language, maximality normally has to be relativized to a contextually relevant domain (e.g. von Fintel 1994).

[2] In the examples here and henceforth I use traditional glosses, to avoid descriptive confusion.

Example in (2) shows a parallel case from Persian.

(2) Dirooz panj ta sag did-am. Emrooz yeki-shoon-**o** did-am.
yesterday five unit dog see.PST-1SG today one-of-them-ACC see.pst-1SG
"Yesterday I saw five dogs. Today I saw one of them." [PERSIAN]

Clearly, the notion of maximality is not appropriate to model the meaning contribution of such determiners. Determiners of similar kinds are known under different terms depending on a particular tradition and their morphological makeup: as (differential) accusative markers in Turkish (Turkic) and Persian (Iranian), as possessive markers in Mari (Finno-Ugric) and Buryat (Mongolian), as demonstrative or definite markers in Mordvin (Finno-Ugric).[3]

The goal of this paper is to show that despite different labels, it is productive to consider them a semantically natural class of morphemes, and that both semantic and pragmatic variation within this class can be captured with a single parameter, namely, the range of relations which can hold between individuals from the denotation of the noun phrase and some contextually specified (group of) individual.[4]

The next section focuses on the empirical patterns and in section 3 I propose a preliminary fomalisation capturing attested semantic variation. In section 4 I probe into pragmatic properties of these determiners, which I then relate them to the semantic variation in section 5. In section 6 I offer brief conclusions.

2 Empirical patterns

As a preliminary descriptive device, I will use a notation from Enç (1991), proposed to capture the use of Turkish differential object markers. In (3), i and j are indices pointing to the referent of the NP in question and some other referent to which the former stands in a subset relation.

(3) Every $[_{NP}\ \alpha]_{<ij>}$ is interpreted as $\alpha(x_i)$:
$x_i \subseteq x_j$ if $NP_{<ij>}$ is plural
$\{x_i\} \subseteq x_j$ if $NP_{<ij>}$ is singular. Enç (1991, 7)

As will be shown below, languages differ with respect to what kind of relation can hold between the referents of i and j. Below, in addition to Turkish, considered here as a baseline case, I focus on the variation within the Finno-Ugric group to which I add data from Buryat, a Mongolian language, mostly based on original fieldwork. This sample seems to exploit nearly all possible pattern combinations and allows to draw a model of semantic typology for this family of determiners.

2.1 Turkish

The accusative marker in Turkish, in addition to the contexts featuring what can be called a proper superset antecedent, as in (1), also appears in contexts which provide an antecedent with which the intended referent of the DP in question can be identified, (4), as well as whenever a suitable antecedent is provided by a local, (5), and a global discourse situation, (5), to use the terms of Hawkins (1991).

[3] As will be shown below, the Mordvin labels are clear misnomers on the Fregean treatment of definites.
[4] Henceforth I will use the term antecedent as an informal shortcut for an individual/group of individuals which verify the use of an antecedent expression.

(4) Oda-m-a bir kız gir-di. Kız-***(ı)** tanı-dı-m.
room-1SG-DAT a girl enter-pst.3sg girl-ACC recognize-PST-1SG
"A girl entered my room. A recognized the girl." [Turkish]

(5) Kapı-***(yı)** kapat!
door-ACC close.IMP.2SG
"Close the door!" [Turkish]

(6) Güneş-***(i)** gör-dü-m.
sun-ACC see-PST-3SG
"I saw the sun." [Turkish]

Turkish also has a series of possessive suffixal determiners, which can only express properly possessive relations and in the object position combine with the accusative suffix, as in (7).

(7) Deniz ev-**in-i** sat-acak.
Deniz house-3SG-ACC sell-FUT.3SG
"Deniz will sell his house." [Turkish]

2.2 Komi

A similar pattern is found in Komi (Finno-Ugric). As most other Finno-Ugric languages, Komi has a paradigm of possessive suffixes. These suffixes attach to a nominal or nominalized stem and encode person and number features of an implicit or explicit expression denoting an individual which stands in some sort of a contextually recoverable relation to elements from the denotation of the NP, as in (8) where the suffix encodes the features (3rd person singular) of the possessor.

(8) Petra-lyn ponm-**ys**
Peter-GEN dog-3SG
"Peter's dog" [Komi][5]

In a number of Finno-Ugric, including Komi, the distribution of the possessive suffixes extends beyond contexts involving an entity-to-another-entity relation (i.e. possessive or genitival relations proper). For instance, a 3rd person singular person suffix appears in contexts parallel to the Turkish case with a proper superset antecedent in (1):

(9) lavka təryt va-i-sny kuim pyzan. ton mi yti pyzan-**#(se)** n'eb-i-m.
store yesterday bring-PRT-3PL three table today we one table-3SG.ACC buy-PRT-1PL
"Yesterday they brought three tables to the store. Today we bought one table." [Komi]

In (10) there is an antecedent identical to the intended referent (cf. Turkish (4)).[6]

(10) me mun-i ul'iča kuz'a i ad'd'-il-i pon. ponm-***(ys)** kuč'-i-s uut-ny.
I walk-PRT street along and see-ITER-PRT dog dog-3SG start-PRT-3 bark-INF
"I was walking down the street and saw a dog. The dog started barking." [Komi], Kashkin (2008)

Finally, as in Turkish, in Komi a possessive determiner appears if the existence of a potential referent is established in the discourse situation, either local, (11), or global, (12).

[5]Komi data are from a Komi Izhem dialect spoken in Muzhi, Shuryshkary district, Yamalo-Nenets region, Russian Federation. Finno-Ugric and Buryat data are presented in transliterations (from Cyrillic) without capital letters.

[6]I use # and * signs to distinguish sharp and mild infelicity respectively. The contrast is empirically noticeable in speakers' judgements and, I suggest, has theoretical grounds, as discussed in section 5.

(11) əbes-***(se)** s'ipt-i!
door-3SG.ACC close-IMP
"Close the door!" [KOMI], Kashkin (2008)

(12) šond'-***(ys)** dzeb-s-i-s.
Sun-3SG dep-DETR-PRT-3SG
'The sun has set.' [KOMI]

2.3 Buryat

In Buryat, the distribution of possessive suffixes covers properly possessive, superset, and identical antecedent contexts, illustrated in (13), (14), and (15) respectively.

(13) gɛr-**ɛn'** exɛ.
house-3SG big
"His house is big." [BURYAT][7]

(14) bi gurban ajaga abaab. nɛgɛ ajag-ii**#(-n')** ɛgɛš-ɛdɛ bɛlɛglɛ-xɛ-b
I three cup bought one cup-CNT-3SG sister-DAT give-POT-1SG
"I bought three cups. One cup I will give to (my) sister." [BURYAT]

(15) manaj tosxondo šɛnɛ gɛr bar'-aa. gɛr(**-ɛn'**) exɛ.
we village new house build-PRT.3SG house-3SG big
"In our village a new house was built. That house is big." [BURYAT]

In the context of a potential referent available in the discourse but not picked up by a linguistic antecedent, the use of the suffixal determiner is infelicitous, (16).

(16) xaxad hʏni hara**(#-n')** gar-aa.
middle night moon-3SG come-PRT.3SG
"The moon came out in the middle of the night." [BURYAT]

Buryat also has differential accusative markers. The latter are used in contexts of direct anaphora, proper superset antecedents, and referents whose existence is guaranteed by a non-linguistic context, as in (17), (18), and (19) respectively.

(17) ʏstɛr ujls-ɛdɛ bi noxoj xar-aa-b. munoodɛr tɛrɛ noxoj**(-e)** ɛdeɛl-ɛɛ-b.
yesterday stree-DAT I god see-PRT-1SG today that dog-ACC feed-PRT-1SG
"Yesterday I saw a dog on the street. Today I fed that dog." [BURYAT]

(18) bi gurban ajaga abaab. nɛgɛ ajagy**-e** ɛgɛš-ɛdɛ bɛlɛglɛ-xɛ-b
I three cup bought one cup-ACC sister-DAT give-POT-1SG
"I bought three cups. One cup I will give to (my) sister." [BURYAT]

(19) munθθdɛr xadaj-n dɛɛrɛ bi nara-**je** xar-aa-b.
today mountain-GEN on I sun-ACC see-PRT-1SG
"I saw the sun above the mountains today." [BURYAT]

2.4 Mari

Mari showcases yet another pattern in terms of the range of contexts in which a suffixal determiner can be used. Along with possessive uses, (20), the only other type of context licensing the use of bound determiners are those with a proper superset antecedent, as in (21).

[7]The Buryat data come Barguzin dialect of Buryat spoken in Baraghan, Buryat Republic, Russian Federation.

(20) üdər-**žö** tud-əm sərəkt-ən.
daughter-3SG he-ACC make.angry-PRT
"His daughter made him angry." [MARI][8]

(21) məj kum kniga-m nal-ən-am. ik kniga**#(-ž)**-əm Kost'a-lan pölekl-em.
I three book-ACC buy-PRT-1SG one book-3SG-ACC Kost'a-DAT give-PRS.1SG
"I bought three books. I will give one of them to Kost'a." [MARI]

Contexts with anaphoric, (22), or situational identity, (23), are excluded in Mari:

(22) Vasja kniga-m nal-ən. Tač'e tudo (tide) kniga-**(*ž)**-əm lud-eš.
Vasja book-ACC buy-NARR.3SG today he that book-(*3SG)-ACC read-PRS.3SG
'Vasja bought a book. Today he is reading that book.' [MARI], Simonenko (2014)

(23) Petər-e-za omsa-**(*žə)**-m!
close-IMP-2SG door-(*3SG)-ACC
'Close the door!' [MARI], Simonenko (2014)

2.5 Mordvin

Finally, a split pattern is found in Mordvin. Possessive determiners cover only properly possessive relations, (24), whereas a suffix a paradigm traditionally labelled "definite" or "demonstrative" and not marking person features appears in all the context where Turkish uses an accusative and Komi a 3rd person singular possessive determiner. (25) illustrates the proper superset antecedent case.

(24) Maša n'ej-əz'ə son' c'or-**ənc**.
Masha meet-PST.3SGO.3SGS his son-3SG.GEN
'Masha met his son.' [MORDVIN][9]

(25) Olɛ rama-s' kolmə kniga-t. fke kniga**#(-t')** son kaz-əz'ə
Ol'a buy-PST.3SG three book-PL one book-DEF.SG.GEN she give-PST.3SGO.3SGS
Kost'-ən'd'i.
Kost'a-DAT
'Ol'a bought three books. She gave one book to Kost'a.' [MORDVIN]

The term "definite" is a misnomer if we reserve it for the cases of iota- or maximal quantification (relativized to a domain): in (25) *fke kniga-t'* cannot be sensibly construed as denoting a maximal individual with the property of being a book. This determiner can also be used in contexts with an anaphoric antecedent, (26) or a situationally accessible referent, (27).

(26) mon ... n'ɛj-ən' pin'ə, i pin'ə***(-s')** uv-əma-n'.
I ... see-PST.1SG dog and dog-DEF.SG bark-PST-1.O-SG.O.3SG.S
'I ... saw a dog, and the dog barked at me.' [MORDVIN], Kashkin (Forthcoming)

(27) t'ɛči ši***(-s')** valdəpt-i valctə.
today sun-DEF.SG shine-NPST.3SG bright.EL
'Today the sun is shining brightly.' [MORDVIN], Kashkin (Forthcoming)

[8]The Mari data are from a dialect of Meadow Mari spoken in Staryj Torjal, Republic Mari El, Russian Federatioin.

[9]Mordvin data are from a Moksha Mordvin dialect of Lesnoe Tsibaevo, Temnikovo region, Mordvin Republic, Russian Federation.

The patterns found in this micro-typological sample are summarized in Table 1 with an extension of Enç's notation.

Pattern		Tur	Tur obj	Kom	Bur	Bur obj	Mar	Mor	Mor "def"
A.	x_i owned by x_j	✓	✗	✓	✓	✗	✓	✓	✗
B-i.	$x_i \subset x_j$ if x_i is pl	✗	✓	✓	✓	✓	✓	✗	✓
B-ii.	$\{x_i\} \subset x_j$ if x_i is sg	✗	✓	✓	✓	✓	✓	✗	✓
C.	$x_i = x_{j_{context}}$	✗	✓	✓	✓	✓	✗	✗	✓
D.	$x_i = x_{j_{disc.sit.}}$	✗	✓	✓	✗	✓	✗	✗	✓

Table 1: Uses of suffixal determiners (possessive paradigm unless indicated otherwise)

3 Semantic variation

We are now in a position to identify parameters of variation, as a first step in developing a unified account of the semantics of suffixal determiners.

The first axis of variation is the (non)acceptability of a suffixal determiner in "anaphoric identity" contexts (Pattern C in table 1). This sets Turkish, Komi, Buryat, and Mordvin ("definite") apart from Mari and Mordvin (possessive). That is, only in the former group the use of a suffixal determiner is felicitous when the intended referent of the relevant DP is identical to the individual verifying the truthful utterance of an antecedent expression.

Another dimension of variation is the (un)acceptability of a suffixal determiner in contexts where an individual to be identified with the intended referent of the relevant DP is given in the discourse situation (rather than in a linguistic context). Along this dimension, Turkish accusative, Komi possessive, Mordvin "definite", and Buryat accusative determiners contrast with Buryat, Mari, and Mordvin possessive determiners (Pattern D).

Finally, there is a contrast between Mordvin, Turkish, and Buryat on the one hand and Komi and Mari on the other with respect to whether there is a designated paradigm for properly possessive uses. In Buryat the split is operative only in the object position, where possessive contexts are covered by a possessive rather than an accusative suffix.

I take the meaning component common to all the determiners considered above to be a relation between an antecedent and elements from the denotation of the head noun. This can be implemented as a relational variable R in the denotations of the bound determiners, adopting the label proposed in Elbourne (2008) for the relational component in the semantics of English demonstratives. The requirement to have an antecedent will be modelled as a silent individual pronoun at the left periphery of DP, which can be either bound or mapped to an individual by a context-dependent assignment function. I will also assume a silent situation pronoun in the structure which fills the Kratzerian situation argument in the denotation of a determiner.

The first and the third points of variation can be formally captured by assigning different ranges to the relational variable R, as in (28).

(28) $[\![det]\!] = \lambda P_{<e,<s,t>>} \,.\, \lambda y_e \,.\, \lambda x_e \,.\, \lambda s_\sigma \,.\, P(x)(s) \,\&\, R(x)(y)$

where R = possession	Mordvin poss, Turkish poss
where R = inclusion, identity	Mordvin def, Buryat obj
where R = possession, inclusion	Mari, Buryat poss
where R = possession, inclusion, identity	Turkish obj, Komi

Having considered the pragmatic aspect of the variation in section 4, I will argue that these ranges form non-accidental clusters of values, although for the moment this may look as a purely descriptive procedure. I will also sketch a solution for capturing the second point of variation.

As a toy example of semantic composition, let us consider Mari form *pij-že*, genuinely ambiguous between "his dog" and "one of those dogs" in (29). The relatum argument is filled by a silent pronoun (with an index i). Although this is not central to the current discussion, I assume that the person features a determiner bears trigger presuppositions about the identity of the potential antecedent. Following common conventions, I implement these as restrictions on the domain of the corresponding argument, which translate into definedness conditions of the resulting function.[10]

(29) $[\![3sg]\!]^{g,c}([\![dog]\!]^{g,c})([\![i]\!]^{g,c})$ is defined if g(i) is not a speaker or hearer,
if defined, $[\![3sg]\!]^{g,c}([\![dog]\!]^{g,c})([\![i]\!]^{g,c}) = \lambda x\ .\ \lambda s\ .$ x is a dog in s and R(x)(g(i)),
where R = possession, inclusion

4 Pragmatic variation

Another dimension of variation, in addition to the range of relations covered by the suffixal determiners in our sample, has to do with the range of interpretations available in negative and intensional contexts. In this respect, determiners considered here fall into two groups, those which are compatible with the both narrow and wide scope existential interpretation and those compatible only with the wide scope interpretation.

Turkish, Mari, and Mordvin possessive determiners belong to the former group. Narrow scope readings are illustrated in (30), (31), and (32) where the existence of an individual with the property denoted by the noun phrase is effectively negated.

(30) Ben-im kız kardeş-**im** yok.
I-GEN.1SG sister-1SG not.exist-3SG
'I don't have (a) sister.' [TURKISH]

(31) myj-yn aka-**m** uke.
I-GEN sister-1SG be.NEG
"I don't have a sister." [MARI]

(32) mon' aš sazər-**əz'ə**
I NEG sister-1SG
"I don't have a sister." [MORDVIN]

In contrast, Mordvin "definite" determiners, (33), and Buryat and Komi possessive determiners, (35) & (37), are only compatible with a wide scope existential interpretation. A narrow scope interpretation is available only if there is no determiner, as (34), (36), and (38) show.

(33) men' vele-sɛ-nɛk aš sel'skəi predsedat'el'-**s**.
we.GEN village-INESS-1PL NEG local head-DEF
"The local head is not in our village." [MORDVIN]

(34) men' vele-sɛ-nɛk aš sel'skəi predsedat'el'.
we.GEN village-INESS-1PL NEG local head
"There is no local head in our village." [MORDVIN]

[10] I assume that if no quantifier is present, existential closure applies to the individual argument.

(35) minii exɛ noxoj**-mni** ugy.
I.GEN big dog-1SG NEG
"My big dog is not here." [BURYAT]

(36) minii exɛ noxoj ugy.
I.GEN big dog NEG
"I don't have a big dog." [BURYAT]

(37) menam abu pon**-me**.
I.GEN NEG dog
"My dog is not with me." [KOMI]

(38) menam abu pon.
I.GEN NEG dog
"I don't have a dog." [KOMI]

For the case of the accusative suffixes in Turkish and Buryat, I probe for the availability of a narrow scope interpretation with respect to intensional predicates, since negation with an existence predicate is not a syntactic option for these markers and in non-intensional contexts the test becomes less reliable because it is more difficult to completely rule out a wide scope interpretation. As (39) an (40) show, in both languages the accusative marker is out with predicates of creation in intensional contexts, which only allow for a narrow scope interpretation.

(39) Kim (bir) ev**(*-i)** yap-mak ist-iyor.
Kim a/one house-acc make-INF want-PROG.3SG
"Kim wants to build a house." [TURKISH]

(40) Bair gɛr(***-e**) barixa hana-taj.
Bair house-ACC build desire-COM.3SG
"Bair wants to build a house." [KOMI]

Table 2 gives a summary of the semantic and pragmatic patterns together.

PATTERN		TUR	TUR OBJ	KOM	BUR	BUR OBJ	MAR	MOR	MOR "DEF"
A.	x_i OWNED BY x_j	✓	✗	✓	✓	✗	✓	✓	✗
B-i.	$x_i \subset x_j$ IF x_i IS PL	✗	✓	✓	✓	✓	✓	✗	✓
B-ii.	$\{x_i\} \subset x_j$ IF x_i IS SG	✗	✓	✓	✓	✓	✓	✗	✓
C.	$x_i = x_{j_{context}}$	✗	✓	✓	✓	✓	✗	✗	✓
D.	$x_i = x_{j_{disc.sit.}}$	✗	✓	✓	✗	✓	✗	✗	✓
E.	NARROW SCOPE	✓	✗	✗	✗	✗	✓	✓	✗

Table 2: Uses of suffixal determiners (possessive paradigm unless indicated otherwise)

5 Deriving the variation

With regard to table 2, notice that there is a perfect negative correlation between an identity relation with a context antecedent being in the range of available relations and the possibility of a narrow scope reading with respect to negation. I propose that this is not an accident. Rather, this patterns follows from the assumption that if R can take an identity relation value, a determiner carries the presupposition that there exists an element (in the relevant domain) with the nominal property in the relation R to the antecedent (cf. Elbourne (2008)'s treatment of demonstratives). One can check that in contexts where there is an antecedent, for the identity relation case this is formally equivalent to the requirement that the antecedent have the nominal property. This requirement is justified given general constraints on the use of anaphoric determiners. For instance, this captures the infelicity of the following anaphoric chain in English (and in any other language I am familiar with, for that matter): *#a pig ... That dog ...*). In other words, this presupposition naturally accompanies identity relations since

otherwise the resulting expression would have been wrongly predicted to hold of individuals which have antecedents without the relevant nominal property. I therefore revise the lexical entry for the determiners which have identity in the range of their relational variable, as in (41).

(41) $[\![det]\!] = \lambda P_{<e,<s,t>>} \,.\, \lambda y_e \,.\, \lambda x_e \,.\, \lambda s_\sigma : \exists x[P(x)(s) \,\&\, R(x)(y)] \,.\, P(x)(s) \,\&\, R(x)(y)$,
where R = ... identity ...

Now if "definite" determiners in Mordvin, possessive determiners in Buryat and Komi, and accusative determiners in Turkish and Buryat (all those with checkmarks in Pattern C line, anaphoric identity relation) trigger existence presupposition, it explains why they are only compatible with a wide scope existential interpretation. A context which satisfies this presupposition is logically incompatible with negating the existence of individuals with the nominal property standing in relation R to the antecedent or with asserting the desirability of their creation. This captures the perfect negative correlation between Patterns C and E.

I propose that this presupposition is also responsible for blocking Turkish and Mordvin possessive suffixes from the contexts with a superset antecedent, which makes them contrast with their Mari counterparts which appear in such contexts. Superset antecedent contexts are different from properly possessive ones in that the existence of a superset entails the existence of its subparts, which, assuming the Maximize Presupposition principle (Heim 1991, Chemla 2008, Singh 2009), gives rise to a grammatical pressure to use a determiner which triggers existence presupposition, which corresponds to an accusative marker in Turkish and a "definite" determiner in Mordvin. The existence of a possessor, on the other hand, in most cases does not entail the existence of a possessee, hence no pressure to use presupposition triggers in possessive contexts.

There also seems to be an empirical contrast in how strongly speakers prefer to use existence presupposition triggers (again, determiners having checkmarks in Pattern C and crosses in Pattern E) in contexts with a direct vs. proper superset antecedent (reflected in the examples with * vs. # signs of unacceptability), the latter contexts more easily allowing for determiner omission. Although this issue will have to await a more thorough investigation, I speculate that since a referent verifying a direct antecedent necessitates the existence of an identical individual (i.e. itself) irrespective of the evaluation situation, the relation between a group and its members is less straightforward and the existential entailment depends on the situation parameter.

Finally, among determiners having an identity relation in the range of their R variable, Turkish and Buryat accusative markers pattern with Komi possessive and Mordvin "definite" determiners in being used in context where the relevant referent does not correspond to a linguistic context (Pattern D in Tables 1–2). In this respect, they contrast with Buryat possessive determiners which require linguistic antecedents. One possible way to model this contrast is by putting different restrictions on the interpretations of the pronominal element in the Logical Forms of these determiners, for instance, by limiting the range of values for the silent pronoun in Buryat to referents already invoked in the preceding discourse. Another possibility is to assume that the Logical Form of the former group, insensitive to the presence of linguistic antecedents, actually does not have either a pronominal element or a relational component in their semantics, and that all the interpretative effects are due to the existential presupposition they trigger. At least for Turkish, an analysis along these lines is proposed by Kelepir (2001). One argument in favour of having a pronominal element in the Logical Form is that suffixal determiners in all these languages are used in noun phrases with an elided noun, as (42) from Buryat illustrates. This pattern is expected assuming that the suffix spells out a pronominal whose antecedent is the same expression as the one that licenses ellipsis.

(42) bi avtobus-abl xožomd-oo-b. hylšɛnxe-ɛr-**ɛn'** jabaa-b.
I bus-abl be.late-PRT-1SG NEXT-INSTR-3SG go-1SG
"I missed the bus. I will take the next one." [BURYAT]

6 Conclusions

Suffixal determiners not associated with maximal quantification have been shown to exhibit a significant degree of variation in their distribution in a sample taken from Finno-Ugric, Monglian, and Turkic languages. I proposed to parametrize the variation by assigning different ranges to the value of the relational component R in the semantics of the determiners. In particular, I argued that the availability of an identity relation as a value for R is always accompanied by an existence presupposition, which, in turn, derives the variation in terms of the availability of narrow scope interpretation in negative and intensional contexts, as well as the patterns of paradigm competition in languages with more than one series of suffixal determiners ("definite" or accusative) series of markers. To the extent that this parametrization is successful, we can talk of a typological class of specific determiners with predictable variation.

References

Chemla, Emmanuel. 2008. An epistemic step for anti-presuppositions. *Journal of Semantics* 25:141–173.

Elbourne, Paul. 2008. Demonstratives as individual concepts. *Linguistics and Philosophy* 31:409–466.

Enç, Mürvet. 1991. The semantics of specificity. *Linguistic Inquiry* 22:1–25.

von Fintel, Kai. 1994. Restrictions on Quantifier Domains. Doctoral Dissertation, University of Massachusetts Amherst.

Hawkins, John A. 1991. On (in)definite articles: Implicatures and (un)grammaticality prediction. *Journal of Linguistics* 27:405–442.

Heim, Irene. 1991. Articles and definiteness. In *Semantics: An International Handbook of Contemporary Research*, ed. Arnim von Stechow and Dieter Wunderlich. Berlin: De Gruyter.

Kashkin, Egor. 2008. Osobennosti upotreblenija posessivnyh pokazatelej v izhemskom dialekte komi-zyrjanskogo jazyka (aspects of the use of possessive markers in izhem komi). In *Acta Linguistica Petropolitana*, ed. N. N. Kazanskij, volume IV, 81–85. Saint Petersburg.

Kashkin, Egor. Forthcoming. Definiteness in moksha. To appear in *Elements of Moksha language in a typological perspective*.

Kelepir, Meltem. 2001. Topics in Turkish syntax: Clausal structure and scope. Doctoral Dissertation, Massachusetts Institute of Technology.

Simonenko, Alexandra. 2014. Microvariation in Finno-Ugric possessive markers. In *Proceedings of the forty third annual meeting of the North East Linguistic Society (NELS 43)*, ed. Hsin-Lun Huang, Ethan Poole, and Amanda Rysling, volume 2, 127–140.

Singh, Raj. 2009. Maximize Presupposition! and Informationally encapsulated implicatures. In *Proceedings of Sinn und Bedeutung*, volume 13, 513–526.

The pragmatics of plural predication: Homogeneity and Non-Maximality within the Rational Speech Act Model*

Benjamin Spector

Institut Jean Nicod, Département d'études cognitives, ENS, EHESS, PSL Research University, CNRS
Paris, France
benjamin.spector@ens.fr

Abstract

This paper offers an account of two puzzling properties of the intepretation of plural definites, known as *Homogeneity* and *Non-Maximality*, within the framework of the Rational Speech Act model.

1 Introduction

Plural definites (PDs) display two well-known properties: *Non-maximality* and *Homogeneity*.

Non-maximality refers to the fact the the quantificational force of plural definites is variable across contexts. While they tend to have universal quantificational force, they easily 'allow for exceptions'. For instance, if there are many windows in a building and if I was asked to make sure that some fresh air enters the building, (1a) below can be judged true if I opened all of the windows but two or three. If the quantifier *all* is added, as in (1b), this is no longer the case.

(1) a. I opened the windows.
b. I opened all the windows.

Homogeneity refers to the fact that even in contexts where a sentence such as (1a) has universal or near-universal force, its negation does not amount to the negation of a universally quantified statement. That is, (2a) does not merely mean that I didn't open all of the windows, but rather suggests that I didn't open *any* of the windows (or at most very few). Again, this effect is removed when *all* is added, as in (2b).

(2) a. I didn't open the windows.
b. I didn't open all the windows.

Križ and Spector offer an account ([8], henceforth KS) of both properties based on two components: a *semantic* component according to which plural predication generates *multiple interpretations*, and a *pragmatic component* that regulates how the resulting underspecified meanings are used and interpreted in different contexts.

The goal of this paper is to reconstruct the pragmatic component within a general framework for pragmatic reasoning, namely the *Rational Speech Act model* ([3]). In so doing, I hope to strengthen the conceptual motivations for our pragmatic component, by showing them to be derivable from general principles of rational conversation, and to make additional predictions regarding how context influences the interpretation of sentences that contain plural definites.

*Thanks to Leon Bergen for useful discussions in relation with this work. The research reported in this work received support from the Agence Nationale de la Recherche (Grants ANR-10-LABX-0087 IEC, ANR-10-IDEX-0001-02 PSL, and ANR-14-CE30-0010- 01 TriLogMean).

2 KS's account

KS's account relies on a specific semantic rule for predicates, and two interpretative principles.

2.1 Semantic component

KS offers a complete compositional semantics whereby, whenever a predicate is applied to a plural object, the resulting denotation consists of a *set of propositions*. Simplifying somewhat, the set of propositions associated with (3a) is given informally in (3b)

(3) Assume that there are exactly three books a, b and c.
   a. Mary read the books
   b. {Mary read $a \vee b \vee c$, Mary read $a \vee b$, Mary read $a \vee c$, Mary read $b \vee c$, Mary read $a \vee (b \wedge c)$, ..., Mary read a, Mary read b, Mary read c, Mary read $a \wedge b$, Mary read $a \wedge c$, Mary read $b \wedge c$, Mary read $a \wedge b \wedge c$ }

This set is obtained by 'plugging' in the position of *the books* all the generalized quantifiers that can be obtained by applying disjunction in all possible ways to all the sums that can be obtained from $\{a, b, c\}$. When negation is applied, as in *Mary didn' read the books*, it applies pointwise to all members of the denotation of the negated sentence or predicate, giving rise to the set {$\neg$(Mary read $a \vee b \vee c$), ..., $\neg$(Mary read $a \wedge b \wedge c$) }.

The members of the denotation of a given sentence are called its *candidate meanings*.

2.2 Interpretative principles

KS's account relies on two interpretative principles

- Ban on overinformative meanings

The first principle states that, among the candidate meanings for S (i.e. all the propositions in the denotation of S), those that are *overinformative* relative to a given Question Under Discussion (QUD) are ruled out. This is what governs the following definition of *relevant candidate meanings*.

1. **Overinformativity relative to QUD**
   Given an underlying QUD Q, modeled as an equivalence relation over possible worlds notated $\sim_Q$, a proposition ϕ is *overinformative* if it makes distinctions between some possible worlds that are equivalent given the QUD. In other words, ϕ is overinformative relative to Q if:
   $\exists w_1, w_2 (w_1 \sim_q w_2 \wedge \phi(w_1) \neq \phi(w_2))$ (where $\phi(w)$ denotes the truth-value of ϕ in w).

2. **Relevant candidate meanings given a QUD**
   A candidate meaning for S is a *relevant candidate meaning* for S relative to a QUD Q if and only if it is not overinformative relative to Q.

Importantly, the QUD with respect to which an utterrance is interpreted need not correspond to an actual interrogative utterance in the prior discourse. Rather, a QUD is simply an equivalence class over worlds that represents what speakers care about at a certain point of a conversation. In fact, there might be uncertainty about what the QUD is, and KS's proposal is based on an idealization. The RSA model that we will develop will not need this idealization.

- Truth on all intepretations.

The second principle states that a sentence S is judged true in the context of a QUD Q only if all its relevant candidate meanings relative to Q are true. This principle entails as a special case the *Stronger Meaning Hypothesis* ([2]). It is possible for both a sentence S and its negation not to be judged true, and so I have implicitly defined a trivalent semantics. In fact, the principle of 'Truth on all interpretations' is similar in spirit to the guiding intuition of supervaluationism.

2.3 Applying of the theory

2.3.1 Capturing homogeneity

Consider the following pair:

(4) a. Mary read the books on her reading list.
b. Mary didn't read the books on her reading list.

Consider first a context where the underlying QUD is something like *Which books did Mary read?*, which corresponds to the equivalence class such that two worlds are equivalent just in case Mary read exactly the same books in both. In this case, none of the candidate meanings for (4a) is overinformative relative to the QUD, because the various candidate meanings differ from each other only with respect to which books Mary needs to have read for them to be true. 'Truth on all readings' then entails that (4a) is judged true if all its candidate meanings are true. One candidate meaning, namely the proposition that Mary read all the books on her reading list, entails all the others, and so (4a) predicted to be judged true if and only if Mary read all the books on her reading list. As to (4b), again no candidate meaning is overinformative relative to the QUD. (4b) is judged true if all candidate meanings are judged true. There is again a candidate meaning that entails all the others, namely the proposition that Mary didn't read any book on the reading list (this proposition is obtained by 'plugging' in the position of *the books on her reading list* the disjunction of all the individual books on the reading list). (4b) is therefore judged true just in case Mary didn't read any books on her reading list.

2.3.2 Non-monotonic contexts

As discussed in KS, 'Truth on all interpretations' captures the intuitive truth-conditions of sentences in which a plural definite is under the scope of a non-monotonic quantifier, as in:

(5) Only one student read the books on the reading list.

In a context where we care about which books each student read, (5) is predicted to be judged true if, for every disjunction of D of pluralities of books on the reading list, only one student read D. These truth-conditions are equivalent to 'there is a student who read all of the books on the reading list and all other students read no books on the reading list' - the conjunction of the two most 'extreme' candidate meanings, namely 'Only one student read all the books' and 'Only one student read at least one of the books'. This appears to be a good result, given, in particular, the experimental results reported in [7].

2.3.3 Capturing non-maximality

Suppose now that Mary has the following reading obligation: 'read at least at least half of the twenty books on her reading list'. Assume further that the underlying question is whether she

in fact did that. In such a context, if we learn that Mary read, say, 15 of the 20 books, it seems that we might truthfully utter something like 'Mary read the books on her reading list, so she will probably pass' (in contrast with 'Mary read all the books on her reading list, so she will probably pass'). This connection between non-maximal readings and the underlying QUD, which is substantiated and discussed at length in [10], is captured as follows. Consider (4a), in the context of the QUD 'Did Mary read at least half of the books on her reading list'. First, note that no candidate meaning entails a negative answer to the question (even the weakest candidate meanings are *compatible* with Mary having read all the books). There is one candidate meaning that is equivalent to the positive answer, namely the proposition that Mary read at least half of the books (which is obtained by taking the disjunction of all pluralities of books that contain at least half of the books). This candidate meaning is not overinformative: it is true in all the worlds in which the answer is positive, false in all others. However, all other candidate meanings are either equivalent to this one, or are overinformative, in the sense that they draw distinctions between worlds in which the answer to the underlying polar question is the same. For instance, the proposition that Mary read, say, at least a specific list of 12 books entails that the answer to the question is positive, but is false in some worlds where the answer is positive. As a result, there is only one relevant candidate meaning, namely the proposition that Mary read at least half of the books, and so (4a) is predicted to convey, in this context, that Mary read at least half of the books on her reading list. In practice this might be too precise a prediction, in that what proposition is exactly conveyed might not be completely clear. In our RSA account, this will be explained by the fact that there might be uncertainty about what the QUD is (cf. section 3.5).

The negative case, (4b), works in the same way. Because the property of being overinformative relative to a QUD is closed under negation, there is again only one relevant candidate meaning for (4b) relative to the QUD 'Did Mary read at least half of the books on her reading list', namely the proposition that Mary didn't read half of the books on her reading list, and so (4b) is predicted, in this context, to convey precisely this proposition.

2.3.4 Motivation

KS argues that their interpretative principles are natural principle of language uses. Ruling out overinformative candidate meanings is rational because it is assumed that speakers respect Grice's maxim of relevance, and so do not provide *irrelevant information*. A proposition that is overinformative relative to a QUD Q does *more* than just addressing Q, and is in this sense not relevant. As to 'Truth on all interpretations', it can be viewed as a rational solution to a coordination problem. Faced with a sentence that can have multiple interpretations that are all relevant and equally plausible in a given context, one has no way to decide which reading is intended. The speaker who uses such a sentence to convey one of its possible meanings to the exclusion of others cannot be sure that the intended reading will be the one that the listener will converge on, and in fact incurs a serious risk that the listener will end up believing something that she herself does not believe, in violation of Grice's maxim of quality. Only if the speaker believes that all the possible meanings are true is this risk eliminated, and so the sentence will typically be used in such cases. The listener can herself reason that the speaker will use the sentence only in such situations, and so the requirement that the sentence be true on all its relevant interpretations for it to be used becomes a convention between speakers and listeners.

One of my goals here is to provide a formally explicit foundation for this reasoning.

3 The Rational Speech Act account

In this section, I attempt to reconstruct the account I have just presented by building on a game-theoretic model of pragmatic inference and message choice, namely the Rational Speech Act model ([3]). For simplicity, I will now assume that sentences with a plural definite are underspecified between just two candidate meanings, i.e. the one where the plural definite has universal force, and the one where it has existential force.

3.1 The basic RSA model.

In the basic RSA model, we start from a *literal listener* L_0 who has a prior probability distribution over worlds and knows the literal meanings of sentences. When hearing an utterance u, L_0 updates her prior distribution by conditionalizing it with the proposition expressed by the literal meaning of u. Then we define a speaker S_1 who wants to communicate her beliefs to L_0 and knows how L_0 interprets sentences (i.e. knows L_0's prior probability distribution and knows that L_0 interprets messages by conditionalization). S_1 is characterized by a utility function U_1 such that the utility of a message u if S_1 believes w is *increasing* with the probability that L_0 assigns to w after updating her distribution with u, and *decreasing* with the *cost* of u. Importantly, if the literal meaning of u is incompatible with w, the utility of u is infinitely negative. S_1's probability of choosing a message u when she wants to communicate w is defined by a function that is increasing with the utility of u relative to w. A parameter λ determines the extent to which S_1 maximizes her utility. Next, we define a more sophisticated listener, L_1, who, when receiving a message u, uses Bayes's rule to update her prior distribution on worlds, under the assumption that the author of u is S_1. A speaker S_2 is then defined exactly like S_1, except that now S_2 assumes that she talks to L_1, not L_0. And so on:

1. The interpretation function $\mathcal{L}$, when applied to a message u and a world w, returns 1 if u is true in w, 0 otherwise.
2. $L_0(w|u) \propto P(w)\mathcal{L}(u,w)$.
3. $U_{n+1}(u|w) = log(L_n(w|u)) - c(u)$
4. $S_{n+1}(u|w) \propto e^{\lambda U_{n+1}(u|w)}$
5. $L_{n+1}(w|u) \propto P(w)S_{n+1}(u|w)$.

3.2 Truth on all readings.

Assume now that some sentences are ambiguous, i.e. we start with a set of multiple interpretation functions $\mathcal{I}$. There are as many different literal listeners as there are interpretation functions. We now define S_1 as believing that she talks to a literal listener L_0 but as being uncertain about which interpretation function L_0 is using. This uncertainty is represented by a probability distribution over $\mathcal{I}$. The rational utility function for such an S_1 is such that the utility of a message u is the *expected utility of* u across interpretation functions. The next listener, L_1, assumes that she is receiving a message from S_1 and does not need to care about which interpretation S_1 is using, since there is no uncertainty about S_1's strategy. Then for the higher levels nothing changes. This leads to:

1. The interpretation function $\mathcal{L}$, when applied to a message u and a world w, returns 1 if u is true in w, 0 otherwise.

2. Update rule for a literal listener who uses an interpretation function $\mathcal{L}$: $L_0(w|u,\mathcal{L}) \propto P(w)\mathcal{L}(u,w)$.
3. $\mathbf{U_1(u|w)} = \sum_{\mathcal{L}\in\mathcal{I}} \mathbf{P}(\mathcal{L}).\mathbf{U_1(u|w,\mathcal{L})} = \sum_{\mathcal{L}\in\mathcal{I}} \mathbf{P}(\mathcal{L_i})[\mathbf{log(L_0(w|u,\mathcal{L}))} - \mathbf{c(u)}]$
4. $S_{n+1}(u|w) \propto e^{\lambda U_{n+1}(u|w)}$ 5. $L_{n+1}(w|u) \propto P(w)S_{n+1}(u|w)$.

Suppose now that, for some interpretation function, the literal meaning of a sentence u is false in w. Then at least one term in the sum in Equation 3 is $-\infty$, and as a result the whole sum is $-\infty$. The utility of u for an S_1 who believes w is thus $-\infty$, and so the probability of choosing u to convey w is 0. 'Truth on all interpretations' is thus derived, and, thereby, homogeneity as well as the behavior of plural definites in non-monotonic contexts.

An important remark is in order. This model significantly departs from the use of *lexical uncertainty* in the RSA literature ([9, 1]), where there are as many S_1's as there are interpretation functions, and it's only at L_1 that reasoning about lexical uncertainty takes place. In these models, the first pragmatic speaker can be viewed as picking an interpretation at random, and is thus less sophisticated than in the model we propose here. The first pragmatic listener then performs Bayesian inference jointly about *both* the speaker's beliefs about the world and the particular reading that the speaker picked. For instance, if one of the reading is very implausible due to the prior probability distribution, the first-level pragmatic listener will assign a low probability to the possibility that the speaker picked that reading. This type of model probably captures important facts about how prior probabilities play a role in disambiguation. The current proposal, on the contrary, cannot capture the role played by prior probabilities in disambiguation. In the next section, I introduce a model that includes questions under discussions, and where the listener, at each level, performs Bayesian inference about the identity of the QUD. This might in principle allow for an *indirect* influence of prior probabilities over worlds on disambiguation, in the following way. If the conveyed meaning of a sentence given a certain QUD is very unlikely (given the priors), the listener will infer that the underlying QUD is a different one, which will in turn have an effect on the interpretation of the sentence. I will not investigate, here, however, the extent to which this indirect mechanism can provide a satisfying theory of disambiguation.

3.3 Non-maximality, QUDs and Overinformativity.

In various works in the RSA framework, speaker utility is relativized to a QUD, which helps account for certain cases of non-literal readings ([6, 5, 9]). E.g., a speaker who is answering 'Are some windows open?' and believes that some but not all are, will receive a utility that is not infinitely negative from using the message *All windows are open*, because this message, though false, entails the correct answer to the QUD. Specifically, the utility of a a message u for a speaker who believes that the actual world is w is no longer determined by the probability that the listener will assign to w after interpreting u, but rather by the probability that the listener will assign to $Q(w)$ after interpreting u, where $Q(w)$ is the true answer to Q in w, i.e. $Q(w) = \{v|v \sim_Q w\}$. This move should thus help us account for the context-sensitivity of plural definites. However, it also carries the risk that *all* will too easily receive a non-universal construal, blurring the contrast between plural definites and quantifiers. In the relevant RSA works, there is uncertainty about the QUD (hence a probability distribution over QUDs), on the part of the listener. The speaker takes into account the listener's uncertainty about QUDs, and this restricts, to a certain degree, the extent of non-literal readings. The basic RSA-model with QUDs is as follows. It assumes that the prior probability distributions over worlds and

QUDs are independent.[1]

1. The interpretation function $\mathcal{L}$, when applied to a message u and a world w, returns 1 if u is true in w, 0 otherwise.
2. Update rule for a literal listener who uses an interpretation function $\mathcal{L}$: $L_0(w|u,\mathcal{L}) \propto P(w)\mathcal{L}(u,w)$.
3. $U_{n+1}(u|w,Q) = log(L_n(Q(w)|u)) - c(u)$
   $= log(\sum_{v \sim_Q w} L_n(w|u)) - c(u)$
4. $S_{n+1}(u|w,Q) \propto e^{\lambda U_{n+1}(u|w,Q)}$
5. $L_{n+1}(w|u) \propto P(w) \sum_Q P(Q).S_{n+1}(u|w,Q)$

Importantly, in this model, the speaker does not care if the listener forms a false belief if this belief is orthogonal to the QUD. Various simulations I have run suggest that this model licenses too easily non-literal readings, in many different cases. For instance, I applied the model to the following case.

1. 3 worlds (no student came, just some of them came, all of them came).
2. Messages: *Some (students came)*, *All*, *Just some*, *No*, *Not all*.
3. Possible QUDs: 'Did some of the students come?', 'Did all of the students come?', 'What's the case?' (= 'Did no, some but not all, or all of the students come?').
4. Message costs: 0 for *Some*, *No*, *All*, 1 for the others.
5. Flat priors on worlds.
6. Priors on QUDs: 0.8 for *Did some of the students came?*, 0.1 for the two others.

The first-level listener (i.e. the first pragmatic listener) assigns a probability of about 0.4 to 'some but not all' after having interpreted *all*, and this probability increases with higher recursion-depth, converging to about 0.47. This appears to be a pretty bad result, since it does not seem to be the case that a sentence such as *All the students came* can be interpreted as being compatible with the possibility that some but not all students came, unless maybe the prior probability distribution over worlds makes it extremely unlikely that all the students came (in which case the sentence might receive a kind of hyperbolic interpretation, suggesting that many students came).[2] Increasing the value of λ does not change the general pattern (the probability of 'some but not all' for a listener who hears *All* remains unacceptably high).

Now, where could a pragmatic ban on overinformative sentences come from? Using *All the students came* to answer *Did some of the students come?* might be misleading in suggesting that the QUD one is answering is actually *Did all students come?*. In the above model, the speaker cares about this only instrumentally, i.e. only to the extent that the listener's wrong belief about what the QUD is might prevent her from identifying the intended meaning of the sentence. I propose to modify the QUD-model so that that the speaker cares not only about communicating the true answer to the QUD, but also about communicating what QUD she is using. This

[1] The independence assumption is simplistic and in general false, since it seems quite clear that the prior distribution over worlds has an influence on which questions conversational participants might be interested in (cf. [11, 4]).

[2] We also need to take care of the potential confound of domain restrictions, which we can by considering sentences such as 'All the students who take my class came'.

amounts to replacing Eq. 3. above with $\mathbf{U_{n+1}(u|w,Q) = log(L_n(Q(w),Q|u)) - c(u)}$ (where $L_n(Q(w),Q|u)$ is the joint probability assigned by the level-n listener to the proposition $Q(w)$ and to the QUD being Q, having heard u). The level-$n+1$ speaker is now modeled as wanting to maximize this joint probability. This gives rise to the following revised model:

1. The interpretation function $\mathcal{L}$, when applied to a message u and a world w, returns 1 if u is true in w, 0 otherwise.
2. Update rule for a literal listener who uses an interpretation function $\mathcal{L}$: $L_0(w,Q|u,\mathcal{L}) \propto P(Q)P(w)\mathcal{L}(u,w)$.
3. $U_{n+1}(u|w,Q) = log(L_n(Q(w),Q|u)) - c(u)$ $= log(\sum_{v \sim_Q w} L_n(w,Q|u)) - c(u)$
4. $S_{n+1}(u|w,Q) \propto e^{\lambda U_{n+1}(u|w,Q)}$
5. $L_{n+1}(w,Q|u) \propto P(w)P(Q).S_{n+1}(u|w,Q)$

Applied to the previous case, with the same parameters, this model yields a reasonable result with enough recursion-depth. The following table reports how the level-5 listener assigns probabilities to worlds (= rows) depending on the message being interpreted (= columns).

	Some	All	Just Some	No	Not All
No	0.00	0.00	0.00	1.00	0.50
Just Some	0.50	0.07	1.00	0.00	0.50
All	0.50	0.93	0.00	0.00	0.00

Even with a strong bias in favor of the question 'Did some of the students come', the sentence *All the students came* ends up assigning a very high probability to the world where all the students came. With further iterations, or higher values for λ, the probability assigned to the 'some-but-not-all' worlds converges to 0.

This new model explains the ban on overinformative sentence in the following way: overinformative sentences can be misleading, even when they are *literally* true: they can mislead the listener's beliefs about which QUD the speaker is trying to answer.

3.4 Full proposal

Now that we have a mechanism to capture 'Truth on all interpretations' and the ban on overinformative sentences, we can put them together, which gives rise to the following model:

1. $L_0(w,Q|u,\mathcal{L}) \propto P(w)P(Q).\mathcal{L}(u,w)$
2. $U_1(u|w,Q) = \sum_{\mathcal{L} \in \mathcal{I}} P(\mathcal{L})(log(L_0(Q(w),Q|u,\mathcal{L})) - c(u))$
3. For $n \geq 1$, $S_n(u|w,Q) \propto e^{\lambda U_n(u|w,Q)}$
4. For $n \geq 1$, $L_n(w,Q|u) \propto P(w)P(Q)S_n(u|w,Q)$
5. $U_{n+1}(u|w,Q) = \ log(L_n(Q(w),Q|u)) - c(u)$

To apply the model to the pair in (4), we need to specify alternative messages. These alternative messages play no role for 'Truth on all interpretations', but they are necessary in order to make sure that overinformative meanings are not chosen, since the reason why an

informative message is not chosen is that is has lower utility than other messages that would be less misleading regarding the nature of the QUD being answered. We also want to check that while the model allows for a flexible interpretation of plural definites, this is not so for quantifiers such as *some* or *all*. I assume that *the books* is less costly than *some (of the) books* and *all (of the) books* because, on top of being less complex when the quantifiers are partitive (*some of the/all of the*), it also has a less complex semantic type than quantifiers. I applied the above model to the following case:

1. 3 worlds as before
2. Messages: *No*, *Some*, *All*, *Not All*, *The*, *Not The*
3. 2 interpretation functions, depending on whether *The* is existential or universal
4. 3 QUDs: 'Is Some?', 'Is All?', 'What's the case?'
5. Cost: c(*The*) = 0<c(*Not The*)=c(*Some*)=c(*All*)=1<c(*No*)=1.5<c(*Not All*)=2.
6. Flat priors on worlds and interpretation functions.
7. 2 types of priors on QUDs: either 'What's the case' or 'Is Some?' has a probability of 0.8, and the two others 0.1.

With a recursion depth of 5 (but for lower values as well) and $\lambda = 5$, the results appear to closely match intuitions. When 'What's the case?' has a prior probability of 0.8, the level-5 listener interprets messages as follows (values are rounded to the second digit):

	Some	All	No	Not All	The	Not The
No	0.00	0.00	1.00	0.50	0.00	0.98
Just Some	1.00	0.00	0.00	0.50	0.09	0.02
All	0.00	1.00	0.00	0.00	0.91	0.00

When 'Is some?' has probability 0.8, we get:

	Some	All	No	Not All	The	Not The
No	0.00	0.00	1.00	0.50	0.00	1.00
Just Some	1.00	0.00	0.00	0.50	0.50	0.00
All	0.00	1.00	0.00	0.00	0.50	0.00

So we capture the context sensitivity of sentences with plural definites, as well as homogeneity. When the salient question is 'What's the case?', the positive sentence in the pair (4) ('Mary read the books on the reading list') receives a universal interpretation, and its negative counterpart means that Mary read no books on the reading list. When the salient question is 'Did Mary read at least some of the books on the reading list?', the plural definite ends up equivalent to an existential quantifier (both for the positive and the negative sentence).

As to quantified sentences, they are not interpreted in such a flexible way, which was the desired outcome – somewhat surprisingly, *some* ends up meaning *some but not all* even when the salient QUD is 'Is Some?', possibly because, as discussed below, it also raises the posterior probability of the question 'What's the case?'.

3.5 Added Benefits

As an added benefit, this model is able to capture the fact that the interpretation of plural definites might appear somewhat vague. This will be the case for instance if the priors over

QUDs are flat. Keeping all the other parameters constant, the level-5-listener then interprets (4a) as conveying that Mary read some of the books and probably read all of them (with probability 0.7).

This model also captures the intuition that utterances can 'raise new questions'. When the underlying question 'Is Some?' has a high prior probability, the messages *Some*, *All* and *Not All* result in the listener now believing that the speaker is answering another question. This is illustrated by the following table, which represents the posterior probability assigned by the level-5 listener to each QUD depending on which message was received, in the case where the QUD with the highest prior probability (0.8) is 'Is Some?'. Whether these specific predictions are empirically adequate is a matter for further research.

	Some	All	No	Not All	The	Not The
Is Some?	0.00	0.00	0.89	0.00	1.00	0.89
Is All?	0.00	0.50	0.00	1.00	0.00	0.00
What's the case?	1.00	0.50	0.11	0.00	0.00	0.11

References

[1] Leon Bergen, Roger Levy, and Noah Goodman. Pragmatic reasoning through semantic inference. *Semantics and Pragmatics*, 9, 2016.

[2] Mary Dalrymple, Makoto Kanazawa, Yookyung Kim, Sam Mchombo, and Stanley Peters. Reciprocal expressions and the concept of reciprocity. *Linguistics and Philosophy*, 21(2):159–210, 1998.

[3] Noah D. Goodman and Andreas Stuhlmüller. Knowledge and implicature: Modeling language understanding as social cognition. *Topics in Cognitive Science*, 5(1):173–184, 2013.

[4] Robert Hawkins and Noah Goodman. Questions and answers in dialogue. 2017.

[5] Justine Kao, Leon Bergen, and Noah Goodman. Formalizing the pragmatics of metaphor understanding. In *Proceedings of the Cognitive Science Society*, volume 36, 2014.

[6] Justine T. Kao, Jean Y. Wu, Leon Bergen, and Noah D. Goodman. Nonliteral understanding of number words. *Proceedings of the National Academy of Sciences USA*, 111(33):12002–7, 2014.

[7] Manuel Križ and Emmanuel Chemla. Two methods to find truth-value gaps and their application to the projection problem of homogeneity. *Natural Language Semantics*, 23(3):205–248, 2015.

[8] Manuel Križ and Benjamin Spector. Interpreting plural predication: Homogeneity and non-maximality. Submitted, 2017.

[9] Daniel Lassiter and Noah D Goodman. Adjectival vagueness in a bayesian model of interpretation. *Synthese*, pages 1–36, 2015.

[10] Sophia Malamud. The meaning of plural definites: A decision-theoretic approach. *Semantics & Pragmatics*, 5:1–58, 2012.

[11] Robert van Rooij. Questioning to resolve decision problems. *Linguistics and Philosophy*, 26:727–763, 2003.

Uniform Definability in Assertability Semantics*

Shane Steinert-Threlkeld

Institute for Logic, Language and Computation, Universiteit van Amsterdam
S.N.M.Steinert-Threlkeld@uva.nl

Abstract

This paper compares two notions of expressive power for a logical language and shows how they come apart. In particular, it introduces a simple framework called *assertability semantics* for handling puzzling features of the interaction of epistemic modals and disjunction. As a consequence of the solution to those puzzles, it is shown that the disjunction is in fact definable: every sentence is equivalent to a sentence without disjunction. But we then prove that the disjunction is not *uniformly* definable: no schematic definition of it can be given in terms of the other connectives of the fragment. We also consider the extension with inquisitive disjunction and prove that it is expressively complete.

As one of its benefits, logical semantics for natural language allows one to precisely answer questions about the *expressive power* of various fragments. Typically, one answers: what classes of structures can be defined by the fragment in question? Because, however, expressions have a given syntactic category and therefore semantic type, logical semanticists should be interested in more fine-grained conceptions of expressive power. In particular, which *operations* on the relevant classes of structures are definable? This paper shows how the two kinds of expressive power can come apart, by studying the interaction of disjunction and modals in the framework of assertability, or state-based, semantics. We show that although the so-called 'split' disjunction allows no new structures to be defined, its operation is not definable, i.e. the connective is not *uniformly* definable.

The paper is structured as follows. Section 1 presents some puzzling data on the behavior of epistemic modals and disjunction. Section 2 introduces a simple assertability semantics and develops some of its basic properties. Section 3 shows that disjunction is definable, while Section 4 introduces the concept of uniform definability and shows that disjunction is not uniformly definable. Section 5 explores the addition of inquisitive disjunction. We show that the resulting system is *expressively complete*, in that every set of states can be defined by a formula. Therefore, disjunction remains definable in this setting. It is conjectured that it still fails to be uniformly definable. Finally, we conclude by discussing future directions. We stress that the paper primarily aims to illustrate the contrast between the two forms of expressive power. The particular assertability semantics developed, while handling some data very elegantly, has empirical problems that are addressed in other work.[1]

1 Some Puzzles of Epistemic Modals and Disjunction

While the primary motivation of the present paper concerns two types of expressive power, the semantic system to be studied is designed to capture some very puzzling behavior of the interaction between (epistemic) modals and disjunctions, which will now be introduced.

*Many thanks to Ivano Ciardelli, Peter Hawke, and three anonymous referees for helpful comments. This work was partially supported by funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013)/ERC Grant Agreement n. STG 716230 CoSaQ.

[1] See Steinert-Threlkeld [2017], chapter 3, "Pragmatic Expressivism and Non-Disjunctive Properties".

Our first puzzle concerns the well-known problem of free-choice possibility, in which certain disjunctions with possibility modals entail conjunctions of possibilities. In particular, in the epistemic case, such inferences appear to arise when disjunctions scope over modals.[2]

(1) Bernie Sanders might or might not win the Democratic nomination.
⇝ Bernie Sanders might win and he might not win.

(2) Maria might be at Science Park or she might be in the city center.
⇝ Maria might be at Science Park and she might be in the center.

These observations motivate (WFC): $\Diamond p \vee \Diamond q$ entails $\Diamond p \wedge \Diamond q$.

Our second puzzle concerns how the interpretation of modals is constrained by their linguistic context, including when embedded under disjunctions. Consider:

(3) Jennifer is at home and might be sick.

(3) is only felicitous if some sick-world is one in which Jennifer is at home. If a speaker knows that Jennifer can't be home sick, it is infelicitous. Dorr and Hawthorne [2013] observe that this 'inherited constraint' survives embedding under disjunction.

(4) Either Jennifer is at home and might be sick, or she's playing hookie.

Again, (4) can only be asserted when it's possible that Jennifer is sick and at home. This observation motivates (IC): $(p \wedge \Diamond q) \vee r$ entails $\Diamond(p \wedge q)$.

For the final puzzle, notice that sentences like (5) and their order variants – so-called *epistemic contradictions* – are notoriously marked.

(5) # It's raining and it might not be.

Moreover, as Yalcin [2007] and others have argued, the markedness survives embedding in a wide variety of contexts, including the antecedents of conditionals and under attitude verbs. This contrasts with the sentences which gloss 'might' as 'for all that the relevant group knows'.

(6) # If it's raining and it might not be, you should take the umbrella.

(7) # José thinks both that it might be raining and it isn't.

Similarly, but more puzzlingly, Mandelkern [2017][3] has observed that disjoining epistemic contradictions also sounds terrible. Suppose that you have a lottery ticket but don't yet know the outcome. Even in such a situation, (8) cannot be felicitously asserted.

(8) # Either I'll win and I might not, or I'll lose and I might not.

Again, the pattern is robust across order variations and different contexts. By contrast with the other embeddings, nearly no existing theory, including dynamic and domain semantics, can capture this infelicity. This discussion motivates (DEC): $(p \wedge \Diamond\neg p) \vee (q \wedge \Diamond\neg q)$ (and their variants) are inconsistent.

2 Assertability Semantics

To handle the phenomena just discussed, we will focus on the language $\mathcal{L}$ that contains proposition letters, negation $\neg$, conjunction $\wedge$, disjunction $\vee$, and a possibility modal $\Diamond$. We write

[2] See, among others, Zimmermann [2000], Geurts [2005], Aloni [2016].
[3] See chapter 1, "Bounded Modality".

$\mathcal{L}_P$ for the language of propositional logic, i.e. without $\Diamond$ and $\mathcal{L}^-$ for the language without $\vee$. Throughout, $\Box\varphi := \neg\Diamond\neg\varphi$.

We will call an *information model* a pair $M = \langle W, V\rangle$ of a set of possible worlds and a valuation V assigning subsets of W to proposition letters. Formulas of $\mathcal{L}$ will be interpreted at *information states* $\mathbf{s} \subseteq W$. We recursively define the relation $M, \mathbf{s} \Vdash \varphi$, to be read as "$\varphi$ is assertable relative to information $\mathbf{s}$". This is intended to capture the following: if an agent has the information $\mathbf{s}$ as her belief-state, then it is epistemically appropriate for her to assert φ.

Definition 1 (Hawke and Steinert-Threlkeld [2016])**.**

$\mathbf{s} \Vdash p$	iff	$\mathbf{s} \subseteq V(p)$
$\mathbf{s} \Vdash \neg\varphi$	iff	for every $w \in \mathbf{s}, \{w\} \nVdash \varphi$
$\mathbf{s} \Vdash \varphi \wedge \psi$	iff	$\mathbf{s} \Vdash \varphi$ and $\mathbf{s} \Vdash \psi$
$\mathbf{s} \Vdash \varphi \vee \psi$	iff	$\mathbf{s}_1 \Vdash \varphi$ and $\mathbf{s}_2 \Vdash \psi$ for some $\mathbf{s}_1, \mathbf{s}_2$ such that $\mathbf{s} = \mathbf{s}_1 \cup \mathbf{s}_2$
$\mathbf{s} \Vdash \Diamond\varphi$	iff	for some $w \in \mathbf{s}, \{w\} \Vdash \varphi$

We write $\Gamma \Vdash \varphi$ iff for every $M, \mathbf{s}$, if $\mathbf{s} \Vdash \gamma$ for every $\gamma \in \Gamma$, then $\mathbf{s} \Vdash \varphi$; and $\varphi \equiv \psi$ iff $\{\varphi\} \Vdash \psi$ and $\{\psi\} \Vdash \varphi$. Let $[\![\varphi]\!]_M = \{\mathbf{s} : M, \mathbf{s} \Vdash \varphi\}$ and $[\![\varphi]\!] = \{\langle M, \mathbf{s}\rangle : \mathbf{s} \in [\![\varphi]\!]_M\}$.

These clauses are rather intuitive: p is assertable relative to some information just in case that information leaves open only p worlds. $\neg\varphi$ is assertable only if the information leaves open no φ worlds. A disjunction is assertable just in case the information is *covered* by a piece of information corresponding to each disjunct.[4] And $\Diamond\varphi$ is assertable just in case a φ possibility is left open by the information.

We first observe that the non-modal fragment behaves classically: a sentence of propositional logic is assertable at a state just when it is clasically true at every world in that state. This quickly enables us to observe that this system satisfies (DEC). In the next section, we will additionally show that the system satisfies both (IC) and (WFC).

Fact 1. For every $\varphi \in \mathcal{L}_P$, (i) $\mathbf{s} \Vdash \varphi$ iff $\{w\} \Vdash \varphi$ for every $w \in \mathbf{s}$; (ii) $\{w\} \Vdash \varphi$ iff $v^*_w(\varphi) = 1$ where v^*_w is the classical propositional extension of the valuation given by $v_w(p) = 1$ iff $w \in V(p)$.

Fact 2. Epistemic contradictions are inconsistent: for every $\varphi \in \mathcal{L}_P, M, \mathbf{s}$: $\mathbf{s} \nVdash \varphi \wedge \Diamond\neg\varphi$.

The restriction to formulas without modals is both essential and justified by the literature, where all of the examples are of that type. In particular, the essential restriction is to formulas which are *flat*, in the sense of part (i) of Fact 1. To see that the restriction is essential, we note that $\Diamond p \wedge \Diamond\neg\Diamond p$ turns out equivalent to $\Diamond p \wedge \Diamond\neg p$, which holds at any information state with both a p and a $\neg p$ world. We find this prediction plausible, but leave its defense to future work.

Corollary 1. The assertability semantics satisfies (DEC): for all $\varphi, \psi \in \mathcal{L}_P$, and every $M, \mathbf{s}$:

$$\mathbf{s} \nVdash (\varphi \wedge \Diamond\neg\varphi) \vee (\psi \wedge \Diamond\neg\psi)$$

Proof. For the disjunction to be assertable at $\mathbf{s}$, there would have to be two sub-states whose union is $\mathbf{s}$, one at which each disjunct is assertable. By Fact 2, no such sub-states exist. □

Before proceeding, we record one important fact about the semantics and one definition, both of which will be used later. We use P as a variable for sets of proposition letters, and P_φ for the set of such letters occurring in φ.

[4] See Simons [2005] for more on (super-)covers and disjunctions. Observe that our definition does not require the sub-states for a disjunct to be non-empty. More on that later.

Proposition 1. *For all $\varphi \in \mathcal{L}$, if $M, s \Vdash \varphi$ and $M', s' \Vdash \varphi$, then $M \sqcup M', s \cup s' \Vdash \varphi$, where $M \sqcup M'$ is disjoint union of models, defined in the obvious way.*

Proof. By induction. We show the disjunction case, leaving the rest to the reader. Suppose $M, \mathbf{s} \Vdash \varphi \vee \psi$ and that $M', \mathbf{s}' \Vdash \varphi$. Then there are $\mathbf{s}_1, \mathbf{s}_2$ such that $\mathbf{s} = \mathbf{s}_1 \cup \mathbf{s}_2$, $\mathbf{s}_1 \Vdash \varphi$, and $\mathbf{s}_2 \Vdash \psi$. And *mutatis mutandis* for $\mathbf{s}'$. Then, by the inductive hypothesis, $\mathbf{s}_1 \cup \mathbf{s}_1' \Vdash \varphi$ and $\mathbf{s}_2 \cup \mathbf{s}_2' \Vdash \psi$. Because $\mathbf{s} \cup \mathbf{s}'$ is itself the union of these two unions, we have that $\mathbf{s} \cup \mathbf{s}' \Vdash \varphi \vee \psi$, as desired. □

Definition 2. $M^P := \langle \mathcal{P}(P), V^P \rangle$ where $V^P(p) = \{X \in \mathcal{P}(P) : p \in X\}$.

3 Definability

In this section, we show that disjunction is definable in terms of the other connectives in the following sense: for every formula in the language including disjunction ($\mathcal{L}$), there is a formula in the language without disjunction ($\mathcal{L}^-$) which is equivalent to it. The proof of this result uses a normal form theorem, which will be the main result of this section. The normal form result also yields immediate proofs that the system satisfies (IC) and (WFC). In the next section, we introduce the concept of *uniform definability* and prove that $\vee$ is not uniformly definable.

Our normal form will show that every formula is equivalent to one of the form $\varphi_b \wedge \Diamond\varphi_a^1 \wedge \cdots \wedge \Diamond\varphi_a^n$, where φ_b and all of the φ_a^i are modal-free formulas. This has an intuitive interpretation: every formula places two types of constraint on an information state: it must *entail* s certain piece of information and it must *be compatible with* certain others.If one thinks of the information state as an asserter's belief worlds, then an assertion of φ expresses a single belief and some *abeliefs*, where an agent abelieves p iff p is compatible with her beliefs. Whence the subscripts b and a. We now make this precise, showing how to simultaneously generate the formula to be believed and the set of formulas to be abelieved.

Definition 3. We simultaneously define two translations $(\cdot)_b : \mathcal{L} \to \mathcal{L}_P$ and $(\cdot)_a : \mathcal{L} \to \mathcal{P}(\mathcal{L}_P)$:

$$
\begin{aligned}
(p)_b &= p & (p)_a &= \varnothing \\
(\neg\varphi)_b &= \neg\left((\varphi)_b \wedge \bigwedge (\varphi)_a\right) & (\neg\varphi)_a &= \varnothing \\
(\varphi \wedge \psi)_b &= (\varphi)_b \wedge (\psi)_b & (\varphi \wedge \psi)_a &= (\varphi)_a \cup (\psi)_a \\
(\varphi \vee \psi)_b &= (\varphi)_b \vee (\psi)_b & (\varphi \vee \psi)_a &= \{(\varphi)_b \wedge \varphi_a : \varphi_a \in (\varphi)_a\} \cup \{(\psi)_b \wedge \psi_a : \psi_a \in (\psi)_a\} \\
(\Diamond\varphi)_b &= \top & (\Diamond\varphi)_a &= \{(\varphi)_b\} \cup (\varphi)_a
\end{aligned}
$$

So: p expresses belief in p. Similarly, $\Diamond\varphi$ expresses a trivial belief, but abelief in the belief expressed by φ as well as all of its abeliefs. $\varphi \wedge \psi$ expresses the belief in the conjunction of the beliefs expressed by φ and by ψ as well as all of the abeliefs expressed by each of φ and ψ. Note that $\neg\varphi$ always expresses only a belief; this resembles the fact that $\neg$ also removes inquisitivity in inquisitive semantics. The clause for disjunction will be illuminated in what follows.

Theorem 1 (Normal Form)**.** *For every $\varphi \in \mathcal{L}$, $\varphi \equiv (\varphi)_b \wedge \bigwedge \{\Diamond\varphi_a : \varphi_a \in (\varphi)_a\}$.*

Proof. By induction. We show only the disjunction case and leave the rest for the reader. So suppose that $\varphi \equiv (\varphi)_b \wedge \bigwedge \{\Diamond\varphi_a : \varphi_a \in (\varphi)_a\}$, and *mutatis mutandis* for ψ. Let $\mathbf{s} \Vdash \varphi \vee \psi$. Then there is an $\mathbf{s}_1 \Vdash (\varphi)_b \wedge \bigwedge \{\Diamond\varphi_a : \varphi_a \in (\varphi)_a\}$ and an $\mathbf{s}_2 \Vdash (\psi)_b \wedge \bigwedge \{\Diamond\psi_a : \psi_a \in (\psi)_a\}$ such that $\mathbf{s} = \mathbf{s}_1 \cup \mathbf{s}_2$. Now, by Fact 1, $(\varphi)_b$ holds at every world in $\mathbf{s}_1$. Since $\Diamond\varphi_a$ holds in $\mathbf{s}_1$, there is a φ_a world. Therefore, there is a $\varphi_a \wedge \varphi_b$ world in $\mathbf{s}_1$, which is also in $\mathbf{s}$. Therefore, $\mathbf{s} \Vdash \Diamond((\varphi)_b \wedge \varphi_a)$

for each φ_a. The same holds for ψ. Again by Fact 1, we also have that $\mathbf{s} \Vdash (\varphi)_b \vee (\psi)_b$, since every world satisfies one of the disjuncts from $\mathcal{L}_P$. So, $\mathbf{s} \Vdash (\varphi \vee \psi)_b \wedge \bigwedge \{\Diamond \chi_a : \chi_a \in (\varphi \vee \psi)_a\}$. The reader can verify that the reasoning above holds in reverse as well. □

Corollary 2. The assertability semantics satisfies (IC): for every $\varphi, \psi, \chi \in \mathcal{L}_P$,

$$(\varphi \wedge \Diamond\psi) \vee \chi \Vdash \Diamond(\varphi \wedge \psi)$$

Proof. Theorem 1 and Definition 3 yield that $(\varphi \wedge \Diamond\psi) \vee \chi$ is equivalent to $(\top \vee \chi) \wedge \Diamond(\varphi \wedge \psi)$, which clearly entails $\Diamond(\varphi \wedge \psi)$. □

Corollary 3. The assertability semantics satisfies (WFC): for every φ, ψ:

$$\Diamond\varphi \vee \Diamond\psi \equiv \Diamond\varphi \wedge \Diamond\psi$$

We observe two striking features of this equivalence. On the one hand, $p \vee q$ does not entail $\Diamond p \wedge \Diamond q$. This is because the sub-state required for p or for q is allowed to be empty.[5] But putting a modal under the disjunction forces each sub-state to be non-empty. This allows our system to preserve classical logic for the propositional fragment while still generating wide-scope free-choice inferences. On the other hand, we do not get narrow-scope free choice as an entailment. In other words: $\Diamond(p \vee q) \nVdash \Diamond p \wedge \Diamond q$. A counter-example: a state with a single $p \wedge \neg q$ world. While this might be seen as a problem, we observe that it can be shown that the standard (recursive) pragmatic explanation of narrow-scope free choice[6] can be re-created in this system. Whether this is a plausible combination of the interaction between scope and free-choice licensing will remain for future work.

Proposition 2. *For every $\varphi \in \mathcal{L}$, there is a $\varphi^* \in \mathcal{L}^-$ such that $\varphi \equiv \varphi^*$.*

Proof. By Theorem 1, $\varphi \equiv (\varphi)_b \wedge \bigwedge \{\Diamond\varphi_a : \varphi_a \in (\varphi)_a\}$. By the two parts of Fact 1, the propositional formulas $(\varphi)_b$ and all of the formulas φ_a can be replaced by disjunction-free formulas while preserving equivalence. □

In other words, Proposition 2 says that $\vee$ is definable in terms of $\{\neg, \wedge, \Diamond\}$. It turns out, however, that for given formulas φ, ψ with a disjunction, the equivalent formulas without the disjunction may bear no resemblance to each other. For example, $p \vee q$ is equivalent to $\neg(\neg p \wedge \neg q)$, while $\Diamond p \vee \Diamond q$ is equivalent to $\Diamond p \wedge \Diamond q$, but not to the corresponding De Morgan formula, since negation removes abeliefs. The question thus arises naturally: is this a defect of the De Morgan formulas, or is something deeper happening? In the next section, we answer that something deeper is going on.

4 Uniform Definability

Having thus shown that $\vee$ is definable, we turn to showing that there is no schematic definition of the connective. The precise concept behind the idea of a schematic definition is the following.

Definition 4 (Uniform Definability)**.** Let $\mathcal{L}_1, \mathcal{L}_2$ be two languages interpreted in the same class of models. An n-ary connective $*$ in $\mathcal{L}_1$ is *uniformly definable* in $\mathcal{L}_2$ iff there is a formula $\varphi_*[p_1, \ldots, p_n] \in \mathcal{L}_2$ such that for all $\psi_1, \ldots, \psi_n \in \mathcal{L}_2$, $*(\psi_1, \ldots, \psi_n) \equiv \varphi_*[p_1/\psi_1, \ldots, p_n/\psi_n]$.

[5] Here we depart from, among others, Aloni [2016]. We welcome this departure, but postpone a full discussion.
[6] See Kratzer and Shimoyama [2002], Fox [2007] among others.

The proof of the main result hinges on the way in which the various connectives interact with formulas that are *upward*-closed, in addition to those that are downward-closed. Note that the modal let us define such sets of sets, which are not definable in standard inquisitive semantics. In fact, this failure of downward-closure (also known as persistence) was also one of the key motivating features of the dynamic approach to epistemic modals.[7]

Definition 5. Let X be a set of sets of worlds. X is downward-closed – X is $\downarrow$ – iff if $s \in X$ and $t \subseteq s$, then $t \in X$. X is upward-closed – X is $\uparrow$ – iff if $s \in X$ and $s \subseteq t$, then $t \in X$. We say that X is netural – X is $\sim$ – iff X is neither $\downarrow$ nor $\uparrow$.

We will call $D := \{\uparrow, \downarrow, \sim\}$ the set of *directions* that a set of sets (or a formula) can have.Variables like d_i will range over this set. For a formula φ, we say that φ is d iff $[\![\varphi]\!]_M$ is d for every information model M. Our proof will show that $\vee$ interacts with upward-closed formulas differently than any formula defined without it. As an illustration, we note a simple fact about the disjunction.

Fact 3. If φ is $\uparrow$, then $\varphi \vee \psi$ is $\uparrow$.

Proof. Suppose φ is $\uparrow$ and let $\mathbf{s}$ be such that $\mathbf{s} \Vdash \varphi \vee \psi$ and let $\mathbf{t} \supseteq \mathbf{s}$. We have that $\mathbf{s}_1 \Vdash \varphi$ and $\mathbf{s}_2 \Vdash \psi$ for some $\mathbf{s}_1, \mathbf{s}_2$ such that $\mathbf{s} = \mathbf{s}_1 \cup \mathbf{s}_2$. Because φ is $\uparrow$, it follows that $\mathbf{s}_1 \cup (\mathbf{t} \setminus \mathbf{s}) \Vdash \varphi$. Since $\mathbf{s}_1 \cup (\mathbf{t} \setminus \mathbf{s}) \cup \mathbf{s}_2 = \mathbf{s} \cup \mathbf{t} \setminus \mathbf{s} = \mathbf{t}$, we have that $\mathbf{t} \Vdash \varphi \vee \psi$. □

Definition 6. Let $\varphi[p_1, \ldots, p_n]$ be a formula.

- φ is *d-enforcing* iff $\varphi[p_1/\psi_1, \ldots p_n/\psi_n]$ is d for every $\psi_1, \ldots, \psi_n$.
- φ is *d-promoting* iff $\varphi[p_1/\psi_1, \ldots, p_n/\psi_n]$ is d if some ψ_i is d.

We say an n-ary connective $*$ is d-enforcing (resp. d-promoting) if $*(p_1, \ldots, p_n)$ is. So, for example: $\Diamond$ is $\uparrow$-enforcing and $\vee$ is $\uparrow$-promoting (this is the content of Fact 3).

Our main result will concern the interaction of being $\uparrow$-promoting and $\uparrow$-enforcing. The key idea will be that it is only the disjunction that allows a formula to 'flip' from being downward-closed to upward-closed whenever an upward-closed formula is substituted. Special care will have to be taken to distinguish the disjunction from atoms, which are also (trivially) $\uparrow$-promoting. Before proceeding, we record a couple of helper facts.

Fact 4. φ is $\uparrow$ if and only if $\varphi_b \equiv \top$

Proof. $\Rightarrow$: suppose $\varphi_b \not\equiv \top$. Then there is a $w \in M^{P_{\varphi_b}}$ such that $\{w\} \nVdash \varphi_b$, i.e. $\{w\} \Vdash \neg\varphi_b$. Then, by construction, $M^{P_{\varphi_b}}, \mathcal{P}(P_{\varphi_b}) \Vdash \Diamond\neg\varphi_b$. So, by the epistemic contradiction Fact 2, $M^{P_{\varphi_b}}, \mathcal{P}(P_{\varphi_b}) \nVdash \varphi_b$. From this, it follows that φ_b is not $\uparrow$ since $\varnothing \Vdash \varphi_b$. The $\Leftarrow$ direction is immediate from Theorem 1. □

Fact 5. $\varphi \wedge \psi$ is $\uparrow$ if and only if φ is $\uparrow$ and ψ is $\uparrow$.

Proof. By Fact 4, $\varphi \wedge \psi$ is $\uparrow$ iff $(\varphi \wedge \psi)_b \equiv \top$. But, by Definition 3, we have $\varphi_b \wedge \psi_b \equiv \top$, which holds iff $\varphi_b \equiv \top$ and $\psi_b \equiv \top$ iff (again by Fact 4) φ is $\uparrow$ and ψ is $\uparrow$. □

Theorem 2. *Every formula $\varphi[p_1, \ldots, p_n]$ in $\mathcal{L}^-$ has the following property:*

(*) If φ is $\uparrow$-promoting, then φ is $\uparrow$-enforcing or uniformly equivalent to a proposition letter.

[7] As began by Veltman [1996] and further developed by many since then.

Proof. We proceed by induction on formulas. Proposition letters clearly satisfy (*), by making the consequent true. So assume that φ_1, φ_2 satisfy (*).

- $\neg\varphi_1$: because $\neg\varphi_1$ is $\downarrow$-enforcing, it is not $\uparrow$-promoting and so trivially satisfies (*).
- $\varphi_1 \wedge \varphi_2$: suppose that $\varphi_1 \wedge \varphi_2$ is $\uparrow$-promoting, i.e. for all $\psi_1, \ldots, \psi_n$, if some ψ_i is $\uparrow$, then $\varphi_1 \wedge \varphi_2[p_1/\psi_1, \ldots, p_n/\psi_n]$ is $\uparrow$. By Fact 5, both $\varphi_1[p_1/\psi_1, \ldots, p_n/\psi_n]$ and $\varphi_2[p_1/\psi_1, \ldots, p_n/\psi_n]$ are also $\uparrow$. In other words, φ_1 and φ_2 are also $\uparrow$-promoting. By the inductive hypothesis, each one is thus either $\uparrow$-enforcing or uniformly equivalent to a proposition letter.

  If both φ_1 and φ_2 are $\uparrow$-enforcing, then so too is $\varphi_1 \wedge \varphi_2$, again by Fact 5.

  If φ_1 and φ_2 are uniformly equivalent to the *same* proposition letter – say, p_i – then we also have $\varphi_1 \wedge \varphi_2$ is uniformly equivalent to p_i.

  The only remaining case is when one conjunct – say, φ_1 – is uniformly equivalent to a proposition letter (without loss of generality, suppose it is p_1), and φ_2 is not uniformly equivalent to that same proposition letter. In this case, we have that $\varphi_1 \wedge \varphi_2$ is not uniformly equivalent to a proposition letter. But, for any $\uparrow$ formula ψ_n, we have that $\varphi_1 \wedge \varphi_2[p_1/p_1, \ldots, p_n/\psi_n]$ is equivalent to $p_1 \wedge \varphi_2[p_1/p_1, \ldots, p_n/\psi_n]$. But this entails p_1, and so is not $\uparrow$. Thus, $\varphi_1 \wedge \varphi_2$ is not $\uparrow$-promoting, contradicting the assumption.

  We have thus shown that $\varphi_1 \wedge \varphi_2$ satisfies (*).
- $\Diamond\varphi_1$: because $\Diamond\varphi_1$ is $\uparrow$-enforcing, it satisfies (*). □

Theorem 3. *$\vee$ is not uniformly definable in $\mathcal{L}^-$.*

Proof. $p \vee q$ is $\uparrow$-promoting, but neither $\uparrow$-enforcing nor uniformly equivalent to a proposition letter. Therefore, the previous Theorem shows that no formula can uniformly define it. □

5 Adding Inquisitive Disjunction

Because the language $\mathcal{L}^-$ is relatively weak, it is worth investigating richer fragments. We now consider the language $\mathcal{L}_{\Cup}$, which adds a new symbol $\Cup$ for inquisitive disjunction. We augment Definition 1 with the clause from inquisitive semantics:

$$\mathbf{s} \Vdash \varphi \Cup \psi \quad \text{iff} \quad \mathbf{s} \Vdash \varphi \text{ or } \mathbf{s} \Vdash \psi$$

In this section, we show that $\vee$ is once again definable, but must only conjecture that it is not uniformly definable. The definability result goes as before: we prove a normal form result which finds, for every formula, an equivalent normal form in which $\vee$ does not occur. The normal form has a pleasant character: every formula is equivalent to an inquisitive disjunction of normal forms in the sense of Theorem 1.

Fact 6. $\neg(\varphi \Cup \psi) \equiv \neg\varphi \wedge \neg\psi$

Proof. $\mathbf{s} \Vdash \neg(\varphi \Cup \psi)$ iff $\{w\} \nVdash \varphi \Cup \psi$ for all $w \in \mathbf{s}$. And $\{w\} \nVdash \varphi \Cup \psi$ iff $\{w\} \nVdash \varphi$ and $\{w\} \nVdash \psi$. This latter condition holds for every $w \in \mathbf{s}$ iff $\mathbf{s} \Vdash \neg\varphi \wedge \neg\psi$. □

Theorem 4 (Normal Form for $\mathcal{L}_{\Cup}$)**.** *Every formula $\varphi \in \mathcal{L}_{\Cup}$ is equivalent to a formula φ^* of the form*

$$\bigCup_i \varphi_0^i \wedge \bigwedge_j \Diamond\varphi_j^i$$

where all of the $\varphi_j^i \in \mathcal{L}_P$.

Proof. The base case – atoms – is trivial; so too is the inquisitive disjunction case. Using the inductive hypothesis that each subformula has a normal form, we handle the rest of the connectives as follows.

- $\neg\varphi$: $\neg\left(\bigvee_i \varphi_0^i \wedge \bigwedge_j \Diamond\varphi_j^i\right)$ is equivalent, by Fact 6, to $\bigwedge_i \neg\left(\varphi_0^i \wedge \bigwedge_j \Diamond\varphi_j^i\right)$. Using Theorem 1 and the corresponding Definition 3, this is equivalent to $\bigwedge_i \neg\left(\varphi_0^i \wedge \bigwedge_j \varphi_j^i\right)$, of the desired form.
- $\varphi \wedge \psi$: note that $\left(\bigvee_i \varphi_0^i \wedge \bigwedge_j \Diamond\varphi_j^i\right) \wedge \left(\bigvee_k \psi_0^k \wedge \bigwedge_\ell \Diamond\psi_\ell^k\right)$ is equivalent to $\bigvee_{i,k} \varphi_0^i \wedge \bigwedge_j \Diamond\varphi_j^i \wedge \psi_0^k \wedge \bigwedge_\ell \Diamond\psi_\ell^k$
- $\varphi \vee \psi$: we have that $\left(\bigvee_i \varphi_0^i \wedge \bigwedge_j \Diamond\varphi_j^i\right) \vee \left(\bigvee_k \psi_0^k \wedge \bigwedge_\ell \Diamond\psi_\ell^k\right)$ holds at information state iff it has two sub-states whose union is the whole state, one of which satisfies the left inquisitive disjunction, one of which satisfies the right inquisitive disjunction. The reader can verify that this is therefore equivalent to $\bigvee_{i,k}\left(\varphi_0^i \wedge \bigwedge_j \Diamond\varphi_j^i\right) \vee \left(\psi_0^k \wedge \bigwedge_\ell \Diamond\psi_\ell^k\right)$ We can then apply Theorem 1 to each inquisitive disjunct, to get an equivalent formula without the $\vee$ scoping over conjunctions, as desired.
- $\Diamond\varphi$: note that $\Diamond$ distributes over inquisitive disjunction, so that $\Diamond\left(\bigvee_i \varphi_0^i \wedge \bigwedge_j \Diamond\varphi_j^i\right)$ is equivalent to $\bigvee_i \Diamond\left(\varphi_0^i \wedge \bigwedge_j \Diamond\varphi_j^i\right)$ which is equivalent to $\bigvee_i \Diamond\left(\varphi_0^i \wedge \bigwedge_j \varphi_j^i\right)$ by Theorem 1. The latter is of the desired form. □

Corollary 4. $\vee$ is definable in $\mathcal{L}_{\text{ᴡ}}^-$.

Proof. As before, $\vee$ can be removed from the $\mathcal{L}_P$ formulas while preserving equivalence. □

Conjecture 1. $\vee$ *is not uniformly definable in* $\mathcal{L}_{\text{ᴡ}}^-$.[8]

We note that $\mathcal{L}_{\text{ᴡ}}$ is genuinely expressively richer, because inquisitive disjunctions are not closed under unions. In fact, we go on to show that $\mathcal{L}_{\text{ᴡ}}$ is in a precise sense *maximally* expressive.

Proposition 3. ᴡ *is not definable in* $\mathcal{L}$. A fortiori*, it is not uniformly definable.*

Proof. $p \,\text{ᴡ}\, q$ is not closed under unions: in $M^{\{p,q\}}$, $\{p\} \Vdash p$, and $\{q\} \Vdash q$ (and so each supports $p \,\text{ᴡ}\, q$ as well), but $\{p,q\} \nVdash p \,\text{ᴡ}\, q$. By Proposition 1, no formula in $\mathcal{L}$ is equivalent to $p \,\text{ᴡ}\, q$. □

Our notion of expressive completeness will run as follows: a language is complete if, for any finite set of proposition letters, any set of sets built out of those atoms can be defined by a formula. To make this precise, we introduce the notion of restricting a model to a set of proposition letters. For a set of letters P, $M\restriction_P$ will identify all worlds in W that agree on all of the proposition letters in P. In that sense, $M\restriction_P$ will contain all and only what M can see concerning the atoms in P.

Definition 7. Let M be an information model and P a set of proposition letters. The *restriction of M to P* – also called the P-reduct of M – is the model $M\restriction_P = \langle W/\equiv_P, V_P\rangle$ where $w \equiv_P w'$ iff for every $p \in P$, $w \in V(p)$ iff $w' \in V(p)$ and $V_P(p) = \{[w]_{\equiv_P} : w \in V(p)\}$. For $\mathbf{s} \subseteq W$, we define $\mathbf{s}\restriction_P := \{[w]_{\equiv_P} : w \in \mathbf{s}\}$. For $X \subseteq \mathcal{P}(W)$, we define $X\restriction_P := \{\mathbf{s}\restriction P : \mathbf{s} \in X\}$.

Definition 8. A language $\mathcal{L}$ is *expressively complete* iff: for every finite set of proposition letters P, M, $X \subseteq \mathcal{P}(W)$, there is a formula $\varphi \in \mathcal{L}$ such that $[\![\varphi]\!]_M\restriction_P = X\restriction_P$.

[8] Compare p. 163 of Ciardelli [2016], where he conjectures that $\vee$ is not uniformly definable in the system InqB, which lacks the modal but has a conditional.

Theorem 5. *$\mathcal{L}_{\mathbb{W}}$ is expressively complete.*

Proof. Let P be a finite set of proposition letters and $\mathbf{s}$ a set of worlds. For w, write $p_w := p$ if $w \in V(p)$ and $\neg p$ otherwise. Let $\varphi_w := \bigwedge_{p \in P} p_w$. Note that $w \equiv_P w'$ if and only if $\varphi_w = \varphi_{w'}$. Observe that $|\mathbf{s}\restriction_P| \leq 2^{|P|}$ which, since P is finite, is finite. The reader can verify that $[\![\bigvee_{[w] \in \mathbf{s}\restriction_P} \varphi_w \wedge \bigwedge_{[w] \in \mathbf{s}\restriction_P} \Diamond \varphi_w]\!]\restriction_P = \{\mathbf{s}\restriction_P\}$ i.e. that the formula defines exactly the set $\mathbf{s}\restriction_P$. We call the formula in brackets above $\varphi_{\mathbf{s}}$. Now, let X be a set of sets of worlds. As before, X must be finite. We then have that $[\![\bigvee_{\mathbf{s} \in X\restriction_P} \varphi_{\mathbf{s}}]\!]\restriction_P = X\restriction_P$ which is of the desired form. □

This result is quite remarkable. It is known that both inquisitive logic and propositional dependence logic are expressively complete for *downward-closed* sets of sets.[9] The above result shows that adding a possibility modal – in a sense the simplest kind of upward-closed formula – to inquisitive semantics turns it into an expressively complete language. To put the point in a slogan: *a little bit of modality goes a long way.*

6 Conclusion

We introduced an assertability semantics to account for some puzzling data concerning the interaction of epistemic modals and disjunction. We proved that even though the disjunction is definable, it is not uniformly definable. This result shows that the standard conception of expressive power does not capture all that is of interest for a natural language semanticist. Even if a formal language can express the truth-conditions for every sentence in a fragment of interest, it can still fail to define all the operations denoted by functional vocabulary of the relevant fragment. Finally, we showed that the disjunction remains definable in the presence of inquisitive disjunction, and conjectured that it is still not uniformly definable. We also showed that adding a modal to inquisitive semantics renders it expressively complete.

Much work remains to be done. First, one would like to settle **Conjecture 1**. $\mathcal{L}_{\mathbb{W}}$ would thus be an expressively complete language which fails to uniformly define a natural connective. Secondly, one could investigate the disjunction in the setting of weak negation: $\mathbf{s} \Vdash -\varphi$ iff $\mathbf{s} \nVdash \varphi$, as first studied in Punčochář [2015]. It can be shown that both $\mathbb{W}$ and $\Diamond$ are uniformly definable (by $-(-p \wedge -q)$ and $--\neg\neg -p$, respectively) and that $\{-, \wedge, \neg\}$ is expressively complete. Nevertheless, we introduce two more conjectures.

Conjecture 2. *$\vee$ is not uniformly definable in $\{-, \wedge, \neg\}$.*

Conjecture 3. *$-$ is not uniformly definable in $\mathcal{L}_{\mathbb{W}}$.*

The latter seems especially plausible, given that $--\varphi \equiv \varphi$ for every φ. Furthermore, does the uniform definability result extend to more complicated semantic settings? For example, Aloni [2016], Steinert-Threlkeld [2017], Roelofsen [2017] all move to a *bilateral* setting – simultaneously defining what we would call assertability and deniability conditions – in order to to solve certain problems.[10] Extending the results of this paper to that setting will be non-trivial. Finally, a thorough investigation of the split disjunction (or close analogues) in both larger fragments and other frameworks would be fruitful.

[9] See Ciardelli [2009] for the former and Yang and Väänänen [2017] for the latter. In particular, $\{\neg, \mathbb{W}\}$ is expressively complete for downward-closed sets. Yang [2017] shows that $\mathbb{W}$ is not uniformly definable in dependence logic.

[10] In the system here, $\neg\neg\Diamond p \equiv p \not\equiv \Diamond p$, which has the consequence that $\neg \Box p \equiv \neg p \not\equiv \Diamond \neg p$. Bilateral systems can fix this defect, among others.

References

Maria Aloni. Free choice disjunction in state-based semantics. In *Logical Aspects of Computational Linguistics*, 2016.

Ivano Ciardelli. *Inquisitive semantics and intermediate logics*. Master's thesis, Universiteit van Amsterdam, 2009.

Ivano Ciardelli. *Questions in Logic*. Phd dissertation, Universiteit van Amsterdam, 2016.

Cian Dorr and John Hawthorne. Embedding Epistemic Modals. *Mind*, 122(488):867–913, 2013. doi: 10.1093/mind/fzt091.

Danny Fox. Free Choice and the Theory of Scalar Implicatures. In Uli Sauerland and Penka Stateva, editors, *Presupposition and Implicature in Compositional Semantics*, pages 71–120. Palgrave Macmillan UK, 2007.

Bart Geurts. Entertaining Alternatives: Disjunctions as Modals. *Natural Language Semantics*, 13:383–410, 2005. doi: 10.1007/s11050-005-2052-7.

Peter Hawke and Shane Steinert-Threlkeld. Informational dynamics of epistemic possibility modals. *Synthese*, 2016. doi: 10.1007/s11229-016-1216-8.

Angelika Kratzer and Junko Shimoyama. Indeterminate Pronouns: The View from Japanese. *The Proceedings of the Third Tokyo Conference on Psycholinguistics*, pages 1–25, 2002.

Matthew Mandelkern. *Coordination in Conversation*. Phd dissertation, Massachusetts Institute of Technology, 2017.

Vít Punčochář. Weak Negation in Inquisitive Semantics. *Journal of Logic, Language and Information*, 24(3):323–355, 2015. doi: 10.1007/s10849-015-9219-2.

Floris Roelofsen. Inquisitive live possibility semantics. In *InqBnB Workshop*, 2017.

Mandy Simons. Dividing things up: The semantics of *or* and the modal/*or* interaction. *Natural Language Semantics*, 13(3):271–316, 2005. doi: 10.1007/s11050-004-2900-7.

Shane Steinert-Threlkeld. *Communication and Computation: New Questions About Compositionality*. Phd dissertation, Stanford University, 2017.

Frank Veltman. Defaults in Update Semantics. *Journal of Philosophical Logic*, 25(3):221–261, 1996. doi: 10.1007/BF00248150.

Seth Yalcin. Epistemic Modals. *Mind*, 116(464):983–1026, 2007. doi: 10.1093/mind/fzm983.

Fan Yang. Uniform Definability in Propositional Dependence Logic. *The Review of Symbolic Logic*, 10(1):65–79, 2017. doi: 10.1017/S1755020316000459.

Fan Yang and Jouko Väänänen. Propositional Team Logics. *Annals of Pure and Applied Logic*, 168(7):1406–1441, 2017. doi: 10.1016/j.apal.2017.01.007.

Thomas Ede Zimmermann. Free Choice Disjunction and Epistemic Possibility. *Natural Language Semantics*, 8(4):255–290, 2000. doi: 10.1023/A:1011255819284.

Additive Presuppositions Are Derived Through Activating Focus Alternatives

Anna Szabolcsi

New York University
as109@nyu.edu

Abstract

The additive presupposition of particles like *too*/*even* is uncontested, but usually stipulated. This paper proposes to derive it based on two properties. (i) *too*/*even* is cross-linguistically focus-sensitive, and (ii) in many languages, *too*/*even* builds negative polarity items and free-choice items as well, often in concert with other particles. (i) is the source of its existential presupposition, and (ii) offers clues regarding how additivity comes about. (i)-(ii) together demand a sparse semantics for *too*/*even*, one that can work with different kinds of alternatives (focus, subdomain, scalar) and invoke suitably different further operators.

1 The plot

The particles *too* and *either* are classically recognized as hard triggers of additive presuppositions, and even as a soft(er) trigger.

(1) (It's not the case that) Bill yawned **too**.
hard presupposition: someone other than Bill yawned

(2) (It's not the case that) Bill didn't yawn **either**.
hard presupposition: someone other than Bill didn't yawn

(3) (It's not the case that) **even** Bill yawned.
hard presupposition: Bill was very unlikely to yawn
soft presupposition: someone other than Bill yawned

The additive presuppositions are typically stipulated, not compositionally derived (Abrusán 2011 is an exception). Recent literature has even ignored them (Chierchia 2013:148) or treated them as part of the assertion (Ahn 2014:29, Gajić 2016). While following Chierchia, Ahn and Gajić in various other respects, this paper attempts to account for the source, shape, and presuppositional nature of the additive component.

The account will be presented in two parts. The particles *too*, *either*, and *even* are focus-sensitive (associate with a focused host). This property seems cross-linguistically stable. It is therefore safe for the analysis to rely on focus alternatives; indeed, the analysis would be missing an important generalization if it did not do that. Section 2 outlines the benefits of working with focus alternatives. (i) Following Geurts & van der Sandt (2004) and Abusch (2010), we recognize that focus is a soft existential presupposition trigger, (ii) we explain why that soft presupposition becomes a hard one in the case of *too* and *either*, but not in the case of *even*, and (iii) we trace the anaphoricity of the additive presupposition to the contextual relevance of focus alternatives.

If we only had to account for the behavior of these three particles, we would be almost done. But there are other pertinent cross-linguistic generalizations that a smaller set of languages makes readily visible. Hungarian, Serbo-Croatian, and Hindi are among them.

Section 3 observes that in those languages, a single particle, to be dubbed TOO, expresses, or participates in expressing, the meanings corresponding to *too*, *even*, and *either*, which are realized by three distinct items in English. But the same particle TOO plays a critical role in building **negative polarity items** and possibly **free-choice items** out of indefinites and lexical expressions. We propose that the contribution of TOO must be fundamentally similar in all the contexts where its presence is critical; triggering an additive presupposition and building NPIs/FCIs are but special cases.

Negative polarity and free-choice items are standardly understood to be disjunctions of subdomain alternatives or scalar alternatives. Fox (2007) and Chierchia (2013) argue that free choice and negative polarity involve the exhaustification of such alternatives. But just like TOO cannot be a specialized additive presupposition trigger, it cannot be a specialized exhaustifier; in Hungarian it clearly acts in concert with other particles that plausibly act as exhaustifiers. TOO must have a sparse semantics. We propose that it seeks out alternatives and activates them: forces them to be figured into meaning, with assistance from other operators.

The additive presupposition, then, should be obtained by a suitable operation on some set of alternatives. In view of Section 2, this should be the set of focus alternatives, at least one of which is presupposed to be true. Section 4 proposes to obtain **additivity** by restricting that set to the alternatives distinct from the prejacent using **recursive exhaustification**, plus **local accommodation** of part of the presupposition. Proceeding this way is intended to replicate the standard construal in a somewhat roundabout, but more generally applicable and thus more explanatory manner. The resulting semantics in each TOO-construction depends on what kind of alternatives and what kind of other operators are involved.

2 Additive presuppositions are grounded in focus alternatives

Focus sensitivity. Sentences like (1-2) can be used in at least two kinds of context. In both contexts, *too*/*either* associates with focus and indicates parallelism, although in the first, focus is narrow (4) and in the second, broad (5). The same holds in languages where the counterpart of *too*/*either* (in Hungarian, *is*/*sem*) always attaches to the phrase that bears intonational prominence and not to the end of the sentence. (Similarly for *even* and *még . . . is*.)

(4) a. Mary yawned. $[\text{BILL}]_F$ yawned, too.
Mari ásított. $[\text{BILL}]_F$ is ásított.

b. Mary didn't yawn. $[\text{BILL}]_F$ didn't yawn, either.
Mari nem ásított. $[\text{BILL}]_F$ sem ásított.

(5) a. Mary was fidgeting. $[\text{BILL yawned}]_F$, too.
Mari fészkelődött. $[\text{BILL is ásított}]_F$.

b. Mary wasn't fidgeting. $[\text{BILL didn't yawn}]_F$, either.
Mari nem fészkelődött. $[\text{BILL sem ásított}]_F$.

In each case, the additive presupposition is that some focus-alternative, not identical to the prejacent, is true. In (4), it is the proposition that someone besides Bill yawned (didn't yawn). In (5), it could be the proposition that besides Bill's (not) yawning, some other sign of boredom was (not) in evidence. In what follows, examples with narrow focus will be used, but the claims carry over to broad focus.

Presuppositionality. Why does the additive component have presuppositional status? The most straightforward answer would be that focus induces an existential presupposition. Unfortunately, this does not go without saying, at least not in English. Rooth (1999) famously argued that no existential presupposition was present in (6B):

(6) *A*: Did anyone win the football pool this week?
B: Probably not, because it's unlikely that [Mary]$_F$ won it, and she's the only one who ever wins.
B′: Probably not, because it's unlikely that it's [Mary]$_F$ who won it, and she's the only one who ever wins.
"In this case, I do find the cleft variant incoherent and contradictory. In contrast, the focus variant is fine. This is an argument against systematically giving focus a semantics of existential presupposition."

Clefts are not well understood; it is not ideal to make a poorly understood construction the gold standard for the existential presupposition. However, for our purposes it suffices if plain focus carries a weaker presupposition than a cleft. That claim has been made both directly about focus, and more generally about constructions involving sets of alternatives.

(7) The Background-Presupposition Rule (Geurts & van der Sandt 2004)
Whenever focusing gives rise to a background $\lambda x.\varphi(x)$, there is a presupposition to the effect that $\lambda x.\varphi(x)$ holds of some individual.

(8) Presupposition triggering from alternatives (Abusch 2010)
Default Constraint L
If a sentence γ is uttered in a context with common ground c, and γ embeds a clause ψ which contributes an alternative set Q, then c is such that the corresponding local context d for ψ entails the disjunction of Q.
For example,
   a. $\big[_\gamma$[if John is in a city], [$_\psi$he is in Syracuse and not Binghamton$_\psi$]
   {$_Q$John is in Syracuse, John is in Binghamton}$_\gamma\big]$
   b. $d = c +$ John is in a city
   c. $\{\text{in}(j,s), \text{in}(j,b)\}$

The Background-Presupposition Rule met with agreement, but how that presupposition projects was not sufficiently clear. The Default Constraint L is on safer grounds: Abusch stresses that it yields soft triggers. Not only can the presupposition be locally accommodated, it can be overridden in discourse. The fact that Abusch attributes the default constraint to alternative sets makes it especially suitable to our purposes. As was anticipated in Section 1, too will be argued to be specifically interested in alternative sets.

Soft vs. hard. In view of the above, focus is a soft presupposition trigger. But *too* is invariably cited in the literature as a hard trigger. How come? *Too* is a functional element whose only mission is to induce an additive presupposition. If that could be canceled in discourse, *too* would be vacuous. A principle proposed in another context sensibly rules that out:

(9) The principle of non-vacuity (Crnič 2011:7) The meaning of a lexical item used in the discourse must affect the meaning of its host sentence (either its truth-conditions or its presuppositions).

The significance of non-vacuity is supported by the fact that the additive presupposition of *even* is softer than that of *too*. Imagine Pooh and friends coming upon a bush of thistles. Eeyore (known to favor thistles) takes a bite but spits it out.

(10) Those thistles must be really prickly! Even Eeyore spit them out!

This may be because the main contribution of *even* is its likelihood presupposition, and so *even* does not become vacuous if its additive presupposition is not satisfied.

The presupposition that one of the focus-alternatives is true is also in place when the particle that associates with focus is *only*. Since *only* negates the alternatives that are not entailed by the prejacent, the existential presupposition will be left to the prejacent to satisfy.

Anaphoricity. On the standard analysis (Heim 1990, Kripke 2009), the presupposition of *too* is anaphoric to some contextually salient, or active, individual or individuals. The present proposal has no special anaphoric component but assumes, with Brasoveanu & Szabolcsi (2013), that the presupposition of *too* is merely existential, although it requires contextual relevance. As a variation on the well-known theme, imagine a circle of dissidents who just received word from one of their number that he successfully made it to the free world. They sit around and sigh,

(11) Now Sam is having dinner in New York, too.

The fact that Sam just joined the ranks of New Yorkers is a source of contextual relevance, even though no particular New Yorkers are salient. Or, with Lincoln (1859),

(12) It is said an Eastern monarch once charged his wise men to invent him a sentence, to be ever in view, and which should be true and appropriate in all times and situations. They presented him the words: *"And this, too, shall pass away."*

Focus-alternatives must be contextually relevant, although often the hearer has to figure out what the relevant set is. We contend that the anaphoric flavor of the presupposition of *too* should be a consequence of the contextual relevance requirement on focus alternatives. The reason why this is somewhat important for our analysis is that we are not going to postulate a dedicated additive *too* that could be endowed with further specific attributes. The empirically attested attributes must come from its basic ingredients.

3 Particle TOO in NPIs and FCIs, cross-linguistically

We have argued that *too*/*either* carries an existential presupposition, because it is focus-sensitive, and focus induces the presupposition that one of the focus-alternatives is true. What *too*/*either* adds to this is **additivity**: some focus-alternative **other than the asserted prejacent** is true. That is not yet accounted for. It could be stipulated in the lexical semantics of the particle, if all its occurrences carried an additive presupposition (possibly, over and above requiring the presence of clause-mate negation, cf. *either*, or carrying a likelihood presupposition, cf. *even*).

In many languages, this is not the case. The goal of this paper is to set the additive presupposition in the context of the broader distribution of the pertinent particles in those languages.[1]

First, in such languages, a single particle expresses, or participates in expressing, the meanings corresponding to *too*, *even*, and *either*. (*Is* is a component of *sem*, and *i* of *ni*.)

[1] (13)-(14) only include a small sample of the relevant data, given that NPIs/FCIs are not in the center of

(13)	Hungarian	Serbo-Croatian	Hindi	English
	Mari **is**	**i** Josip	Raam **bhii**	X too
	még Mari **is**	(čak) **i** Josip	Raam **bhii**	even X
	Mari **sem**	**ni** Josip	Raam **bhii**	X either

Second, that same particle also participates in building negative polarity items and free-choice items out of wh-indefinites and lexical expressions.

(14)	Hungarian	Serbo-Croatian	Hindi	English
	valaki **is**	**i**-(t)ko / [bilo (t)ko]	koii **bhii**	anyone, NPI
	még/akár csak Mari **is**	(čak/makar) **i** Josip	Raam **bhii**	even X, NPI
	akár Mari **is**	(čak) **i** Josip	(koii **bhii**)	even X, FCI

The particles *is*, *i*, and *bhii* will be generically referred to as TOO in small caps.

The interest of these observations is that we have well-established theories of how negative polarity and free choice work. NPIs and FCIs are understood to be disjunctions/existentials over subdomain alternatives or scalar alternatives. Furthermore, Chierchia (2013) and Fox (2007) argue that NPIs and FCIs involve the exhaustification of such alternatives.

On Chierchia's theory (in the spirit of Lahiri 1998), an NPI is an existential situated at the low end of a scale (either inherently or via scale truncation). It has obligatorily active (grammaticized) alternatives that must be figured into meaning by exhaustification. Exhaustification leads to a contradiction, unless a decreasing operator is present right below the exhaustifier. That is, the necessity for the NPI to be in a locally decreasing environment is not directly stipulated. We adopt this theory, because it fits nicely with the way Hungarian productively builds NPIs, as will be demonstrated informally.

Consider *még/akár csak Mari **is*** from (14). It is plainly ungrammatical in an upward monotonic environment, i.e. it is an NPI.

(15) Kevesen/*Sokan gratuláltak még/akár csak Marinak is.
'Few/*Many people congratulated even Mari (let alone others)'

Let us proceed item by item. (i) Abrusán (2007) argued that *még* and *akár* are *even*-style exhaustifiers; see also Crnič (2011). (ii) Unlike the indefinite *valaki*, *Mari* does not inherently fall at the low end of any scale. The presence of *csak* brings that about; here *csak* is similar to Dutch *slechts* 'mere(ly)'. Szabolcsi (1994) showed that *csak* can be added to numerals that are downward monotonic ('fewer than n') or non-monotonic ('between n and m' or focused 'n' interpreted as 'exactly n'), but not to irrevocably upward monotonic ones ('more than n').

(iii) What is the particle *is* doing? Notice that *is* is absolutely critical here. *Valaki*, by itself, is 'someone', a PPI, not an NPI. *Még/akár(csak) Mari*, by itself, is a word salad. Chierchia assumes that it is a lexical property of NPIs that they have obligatorily active alternatives. We interpret the Hungarian data as suggesting that activating alternatives is a function that can be delegated to a separate morpheme.

Free-choice items are likewise productively built with particle *is*. Consider *akár Mari **is*** (here *akár* does not alternate with scalar *még*, and the low-end marker *csak* cannot be added).

(16) Akár Mari **is** nyerhet/*nyer.
'Anyone can win/*wins; to pick an arbitrary example, Mari'

this paper. Some comments. The Hungarian and Serbo-Croatian NPIs in (14) are weak (do not occur with clause-mate negation); the negative concord counterparts involve *sem/ni*, cf. *either* in (13). Hindi does not distinguish between weak NPIs and NCIs. The Serbo-Croatian data come from Progovac (1993) and J. Gajić (p.c.), and the Hindi data from Lahiri (1998) and V. Dayal (p.c.).

Fox (2007) proposes that free choice is the result of recursive exhaustification of a set of alternatives with an existential modal under the exhaustifier; Chierchia (2013) recasts that as exhaustification with respect to a pre-exhaustified set of alternatives.

In summary, the following hypothesis seems plausible:

(17) **Too seeks out and activates a set of alternatives**
Hungarian ***is*** (cross-linguistic TOO) seeks out a set of alternatives induced by its host and activates them, so they must be figured into the meaning of the sentence, e.g., by exhaustification. But ***is*** itself does not exhaustify alternatives; it co-occurs with other particles that probably do just that.

What kind of alternatives does TOO recruit, and what kind of semantic operation does it invoke? That depends on what alternatives are available in the given construction, and what kind of operation suits those. The proposal is that TOO is underspecified in this regard. It is possible that the options could be narrowed, but we will not undertake that here. In the realm of negative polarity and free choice, subdomain alternatives or scalar alternatives present themselves. In the realm of expressions with an additive presupposition, we have argued that focus-alternatives present themselves.

4 Too and the additive presupposition

Section 2 observed that *too* and *is* associate with focus, and proposed that presuppositions triggered from focus-alternatives explain the presuppositional character of the additive component. Focus induces a set of propositional alternatives, type $\langle\langle s,t\rangle,t\rangle$, as per Rooth (1992). A set of propositional alternatives is nothing else than the disjunction (join) of the member alternatives: $\{\{w:\varphi_w\},\{w:\psi_w\},\{w:\chi_w\}\} = \{\{w:\varphi_w\}\}\cup\{\{w:\psi_w\}\}\cup\{\{w:\chi_w\}\}$. This puts the focus-alternative set on a par with the core $\exists/\vee$ semantics of NPIs and FCIs.

(18) BILL ásított 'BILL yawned'
assertion: $\text{yawn}_{w^\star}(b)$
focus-alternatives, $ALT: \{\{w:\text{yawn}_w(b)\},\{w:\text{yawn}_w(m)\},\{w:\text{yawn}_w(k)\}\}$
presupposition: $\exists p\in ALT: p_{w^\star}$

(19) BILL nem ásított 'BILL didn't yawn'
assertion: $\neg\text{yawn}_{w^\star}(b)$
focus-alternatives, $ALT: \{\{w:\neg\text{yawn}_w(b)\},\{w:\neg\text{yawn}_w(m)\},\{w:\neg\text{yawn}_w(k)\}\}$
presupposition: $\exists p\in ALT: p_{w^\star}$

In this context, the presence of TOO modifies the presupposition that at least one focus-alternative is true to the effect that at least one focus-alternative **other than the prejacent** is true; this is additivity. TOO plays its role by seeking out the set of focus-alternatives, ALT and relies on some operation that removes the prejacent from ALT, in one way or another. Preliminarily stating this directly in terms of set-theoretic difference,

(20) BILL **is** ásított 'BILL yawned **too**'
assertion: $\text{yawn}_{w^\star}(b)$
$ALT^{DIFF} = \{\{w:\text{yawn}_w(b)\},\{w:\text{yawn}_w(m)\},\{w:\text{yawn}_w(k)\}\}\setminus\{\{w:\text{yawn}_w(b)\}\}$
$= \{\{w:\text{yawn}_w(m)\},\{w:\text{yawn}_w(k)\}\}$
presupposition: $\exists p\in ALT^{DIFF}: p_{w^\star}$

(21) BILL **sem** ásított 'BILL didn't yawn **either**'
assertion: $\neg\text{yawn}_{w^\star}(b)$
$ALT^{DIFF} = \big\{\{w : \neg\text{yawn}_w(b)\}, \{w : \neg\text{yawn}_w(m)\}, \{w : \neg\text{yawn}_w(k)\}\big\}\backslash\{\{w : \neg\text{yawn}_w(b)\}\}$
$= \big\{\{w : \neg\text{yawn}_w(m)\}, \{w : \neg\text{yawn}_w(k)\}\big\}$
presupposition: $\exists p \in ALT^{DIFF} : p_{w^\star}$

In what follows we experiment with a procedure where a version of exhaustification helps produce ALT^{DIFF}. Using such a procedure is desirable because, if viable, it unifies the ways in which the activated alternatives can be figured into the meaning of the sentence, while leaving it open what kind of exhaustification is suitable in each case.

There is a recent line of research that derives conjunctive meanings from disjunctive ones by recursive exhaustification without negating a stronger, conjunctive alternative; a modification of Fox (2007). See Bar-Lev & Margulis (2014) for Modern Hebrew *kol*, Mitrović (2014) for Japanese *mo*, Bowler (2014) for Warlpiri *manu*, Singh et al. (2016) for Child English *or*, and Wong (2017) for Malay *pun*. Of these authors, Mitrović addresses *mo* as an additive particle; his proposal is our closest model. But Mitrović assumes that *mo* itself is a recursive exhaustifier and stipulates presuppositionality, which we cannot literally follow.

A critical assumption is that in the calculation of exhaustification, the disjunction has only subdomain alternatives (the disjuncts) but no scalar, i.e. stronger alternative (the conjunction), and so no conjunctive alternative is negated. Several of the authors justify that with reference to the fact that the given language has no separate word for conjunction, or (in the case of child language) the speaker cannot access that word. For how such recursive exhaustification yields a conjunction, (22-23) replicate Bar-Lev & Margulis (2014).

(22) $\text{EX}(Alt(p))(p)(w) \Leftrightarrow p$ is true in w, and every excludable alternative of p is false in w.
$Excludable(p, Alt(p)) \Leftrightarrow \cap\big\{Alt(p)' \subseteq Alt(p) : Alt(p)'$ is a maximal set in $Alt(p)$ such that $\{p\} \cup \{\neg q : q \in Alt(p)'\}$ is consistent$\big\}$

(23) EX $\text{EX}(a \vee b) = a \wedge b$
$Alt(a \vee b) = \{a \vee b, a, b\}$ Note: $a \wedge b$ is not an alternative.
$\text{EX}_{Alt(a\vee b)}(a \vee b) = a \vee b$ $\{a \vee b, \neg a\}$ and $\{a \vee b, \neg b\}$ are both consistent sets and maximal as such. But $a, b \notin \{a \vee b, \neg a\} \cap \{a \vee b, \neg b\}$.

$Alt(EX_{Alt(a\vee b)}[a \vee b]) = \big\{\text{EX}_{Alt(a\vee b)}[a \vee b], \text{EX}_{Alt(a\vee b)}[a], \text{EX}_{Alt(a\vee b)}[b]\big\}$
$= \big\{a \vee b, a \wedge \neg b, b \wedge \neg a\big\}$

$\text{EX}_{Alt(EX_{Alt(a\vee b)}[a\vee b])}\big[\text{EX}_{Alt(a\vee b)}[a \vee b]\big] =$
$\text{EX}_{\{a\vee b, a\wedge\neg b, b\wedge\neg a\}}[a \vee b] =$
$a \vee b \wedge \neg(a \wedge \neg b) \wedge \neg(b \wedge \neg a) =$
$a \vee b \wedge (a \rightarrow b) \wedge (b \rightarrow a) =$
$a \vee b \wedge (a \leftrightarrow b) = \boldsymbol{a \wedge b}$
Now $a \wedge \neg b$ and $b \wedge \neg a$ are negated; the negations are consistent with $a \vee b$.

Here is a way to produce the same outcome as ALT^{DIFF}, using exhaustification. We stipulate that TOO "bifurcates" the alternative-set into two big alternatives: the prejacent and a flattened-out disjunction of the other alternatives. (All focus-sensitive particles distinguish the prejacent, although not in this same way.) Call the result *BI-ALT*. With *BI-ALT*, the presupposition would be that the prejacent is true or some other alternative is true. But, as per (17), TOO forces the exhaustification of *BI-ALT*; this time recursively, without a scalar alternative. Note that no lexical element serves as a primitive additive particle.

(24) BILL **is** ásított 'BILL yawned **too**'
assertion: $\text{yawn}_{w^\star}(b)$
$BI\text{-}ALT = \{\{w : \text{yawn}_w(b)\}, \{w : \text{yawn}_w(m) \vee \text{yawn}_w(k)\}\}$
EX EX$(BI\text{-}ALT) = \{w : \text{yawn}_w(b)\} \cap \{w : \text{yawn}_w(m) \vee \text{yawn}_w(k)\}$
presupposition: $\exists p \in$ EX EX$(BI\text{-}ALT) : p_{w^\star}$

(25) BILL **sem** ásított 'BILL didn't yawn **either**'
assertion: $\neg\text{yawn}_{w^\star}(b)$
$BI\text{-}ALT = \{\{w : \neg\text{yawn}_w(b)\}, \{w : \neg\text{yawn}_w(m) \vee \neg\text{yawn}_w(k)\}\}$
EX EX$(BI\text{-}ALT) = \{w : \neg\text{yawn}_w(b)\} \cap \{w : \neg\text{yawn}_w(m) \vee \neg\text{yawn}_w(k)\}$
presupposition: $\exists p \in ALT^{DIFF} : p_{w^\star}$

Now (24) presupposes that Bill yawned **and** someone else yawned. The conjunct corresponding to the prejacent can be eliminated; and it must be eliminated under extra-clausal negation:

(26) Nem igaz, hogy BILL is ásított. 'It is not true that BILL yawned too.'

M. Esipova (2017; p.c.) suggests that treating the same content as both at-issue (asserted) and not-at-issue (presupposed) is odd on the global level. Oddness or contradiction may motivate the local accommodation of the prejacent part of the presupposition generated by TOO. The above combination of *BI-ALT* plus local accommodation has some ad hoc elements; it can be hopefully improved upon in future work.

5 TOO in the MO family

In this paper, we have motivated the need for a unified treatment of the various uses of TOO, and made some headway with meeting the challenge. Szabolcsi (2015) proposed that unary TOO belongs to the larger family of "MO particles", which also participate in reiterated constructions with a distributive conjunction interpretation and build universal quantifiers.

(27)

Japanese		Hungarian	
A mo	'A too, even A'	A is	
A mo B mo	'A as well as B'	A is B is	/ mind A mind B
dare-mo	'everyone, anyone'		mind-en-ki

The present claims are entirely compatible with that proposal. They provide the backstory of how TOO/MO particles come to mean what they mean. On the other hand, Szabolcsi (2017) takes up the distinction between Hungarian *is* and *mind*, and argues that morpho-syntactically, there is no unbroken line from the unary particle to the quantifier.

Acknowledgements

Thank-you to M. Esipova and to R. Balusu, L. Champollion, V. Dayal, J. Gajić, D. Lassiter, H. Li, M. Mitrović, C. Roberts, B. Slade, participants of the New York Philosophy of Language Workshop, and the AC 2017 reviewers for discussion, suggestions, and objections.

References

Abrusán, M. (2007). *Even* and free choice *any* in hungarian. In *Proceedings of Sinn und Bedeutung*, volume 11, pages 1–15, http://semanticsarchive.net/Archive/TVkNTE2O/sub11proc.pdf.

Abrusán, M. (2011). Predicting the presuppositions of soft triggers. *Linguistics and philosophy*, 34(6):491–535.

Abusch, D. (2010). Presupposition triggering from alternatives. *Journal of Semantics*, 27:37–80.

Ahn, D. (2014). The semantics of additive *either*. In *Proceedings of Sinn und Bedeutung*, volume 19, pages 20–36, http://semanticsarchive.net/Archive/TVlN2I2Z/sub19proc.pdf.

Bar-Lev, M. and Margulis, D. (2014). Hebrew kol: a universal quantifier as an undercover existential. In *Proceedings of Sinn und Bedeutung*, volume 18, pages 60–76, http://semanticsarchive.net/Archive/jQ5MDU4N/index.html.

Bowler, M. (2014). Conjunction and disjunction in a language without'and'. In *Semantics and Linguistic Theory*, volume 24, pages 137–155.

Brasoveanu, A. and Szabolcsi, A. (2013). Presuppositional too, postsuppositional too. In Aloni, M., Franke, M., and Roelofsen, F., editors, *The dynamic, inquisitive, and visionary life of φ, ?φ, and $\Diamond\varphi$: a festschrift for Jeroen Groenendijk, Martin Stokhof, and Frank Veltman*. ILLC Publications, http://www.illc.uva.nl/Festschrift-JMF/.

Chierchia, G. (2013). *Logic in grammar: Polarity, free choice, and intervention*. Oxford University Press.

Crnič, L. (2011). On the meaning and distribution of concessive scalar particles. *North Eastern Linguistics Society (NELS)*, 41:1–14.

Esipova, M. (2017). Presuppositions under contrastive focus: Standard triggers and co-speech gestures. Ms., New York University. http://ling.auf.net/lingbuzz/003285.

Fox, D. (2007). Free choice disjunction and the theory of scalar implicatures. In Sauerland, U. and Stateva, P., editors, *Presupposition and implicature in compositional semantics*, pages 71–120. Palgrave Macmillan, Basingstoke.

Gajić, J. (2016). Coordination and focus particles (re?)united. In *Proceedings of Sinn und Bedeutung 21, https://sites.google.com/site/sinnundbedeutung21/proceedings-preprints*.

Geurts, B. and van der Sandt, R. A. (2004). Interpreting focus. *Theoretical Linguistics*, 30:1–44.

Heim, I. (1990). Presupposition projection. In *Reader for the nijmegen workshop on presupposition, lexical meaning, and discourse processes. University of Nijmegen*.

Kripke, S. A. (2009). Presupposition and anaphora: Remarks on the formulation of the projection problem. *Linguistic Inquiry*, 40(3):367–386.

Lahiri, U. (1998). Focus and negative polarity in Hindi. *Natural language semantics*, 6(1):57–123.

Lincoln, A. (1859). Address before the wisconsin state agricultural society. http://www.abrahamlincolnonline.org/lincoln/speeches/fair.htm.

Mitrović, M. (2015). *Morphosyntactic atoms of propositional logic:(a philo-logical programme)*. PhD thesis, University of Cambridge.

Progovac, L. (1993). Negative polarity: Entailment and binding. *Linguistics and Philosophy*, 16(2):149–180.

Rooth, M. (1992). A theory of Focus interpretation. *Natural Language Semantics*, 1:75–116.

Rooth, M. (1999). 12 association with focus or association with presupposition? *Focus: Linguistic, cognitive, and computational perspectives*, page 232.

Singh, R., Wexler, K., Astle-Rahim, A., Kamawar, D., and Fox, D. (2016). Children interpret disjunction as conjunction: Consequences for theories of implicature and child development. *Natural Language Semantics*, 24(4):305–352.

Szabolcsi, A. (1994). All quantifiers are not equal: The case for focus. *Acta Linguistica Hungarica*, 42(3-4):171–187.

Szabolcsi, A. (2015). What do quantifier particles do? *Linguistics and Philosophy*, 38(2):159–204.

Szabolcsi, A. (2017). Two types of quantifier particles: quantifier-phrase internal vs. heads on the clausal spine. http://www.nyu.edu/projects/szabolcsi/szabolcsi_research.html.

Wong, D. J. M. (2017). Negative polarity items in malay: An exhaustification account. In Erlewine, M. Y., editor, *Proceedings of GLOW in Asia XI*, Boston. MIT Working Papers in Linguistics, https://deborahjmwong.wixsite.com/site/research.

Quantifiers and verification strategies: connecting the dots

Natalia Talmina[1], Arnold Kochari[2], and Jakub Szymanik[2*]

[1] Johns Hopkins University, Baltimore, Maryland, U.S.A.
talmina@jhu.edu
[2] ILLC, Universiteit van Amsterdam, Amsterdam, the Netherlands
a.kochari@uva.nl
jakub.szymanik@gmail.com

Abstract

In this paper, we replicate the influential study of Hackl (2009), making more specific algorithmic-level predictions based on Hackl's findings. Hackl argued that two semantically equivalent quantifiers *more than half* and *most* are associated with different verification strategies. The results of our experiment diverge in several respects from the original study. We explain the results by focusing on two potential confounds in Hackl's 2009 experimental set-up: different roles that working memory can play in the verification of different quantifiers and individual differences suggesting the use of various cognitive strategies.

1 Introduction

In his influential paper, Hackl (2009) explores whether there is a cognitively significant difference in the specifications of truth conditions of quantifiers *most* and *more than half*, captured below.

(1) a. $[\![\text{most}]\!] = |A \cup B| > |A - B|$
b. $[\![\text{most}]\!] = |A \cup B| > \frac{1}{2}|A|$

(2) a. $[\![\text{more than half}]\!] = |A \cup B| > |A - B|$
b. $[\![\text{more than half}]\!] = |A \cup B| > \frac{1}{2}|A|$

Hackl argues on conceptual and linguistic grounds that (1a) is the preferred option over (1b) for *most*, while (2b) is a better way to express *more than half* compared to (2a). Subsequently, he also suggested that although the two denotations are truth-conditionally equivalent, the way in which they are specified appears to point to distinct verification procedures. *More than half* explicitly calls for dividing the total number of A's in half in the course of verification, while verifying *most* requires comparing the total number of A's that are B's (e.g. the number of dots that are blue) with the number of A's that are not B's (e.g. the number of dots that are not blue).

In order to understand whether there is a difference in verification profiles that are triggered by *most* and *more than half*, Hackl conducted an experiment where participants had to verify visual scenes (pictures containing rows of dots of different colors) against sentences like *Most of the dots are blue* or *More than half of the dots are blue*. He applied the Self-Paced Counting paradigm, which is similar in spirit to the widely used self-paced reading paradigm: instead of having access to the whole scene at once, participants have to press a button to proceed through the scene step-by-step while the time they spend on each screen is measured. Based

*The author have received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007–2013)/ERC Grant Agreement n. STG 716230 CoSaQ.

on the time spent on each screen, one can make inferences about the processes that took place at that point.

The results showed there was no significant difference in overall reaction (i.e. screen inspection) times or accuracy, which Hackl takes to indicate that subjects treated *most* and *more than half* as equivalent expressions. However, there was a significant difference in reaction times *per screen*, when excluding the final screen where the decision was made: verifying *more than half* took subjects consistently longer than verifying *most* (during the inspection of screens which did not yet reveal the correct answer). Hackl observes that this difference makes sense if *most* favors a kind of lead counting strategy — "keeping track of whether the target color leads overall and by how much" (p. 89). Contrary to verifying *more than half*, the lead counting algorithm does not call for dividing the total number of dots in half. The design of the experiment made the task easier for the strategy assumed to be adapted for *most*: in each screen, it was easy to evaluate whether there were more dots in the target color than in the other color, whereas division by half was not as easy in this set-up.

Additional evidence for a direct relationship between meaning and verification comes from other experimental paradigms: Pietroski et al. (2009) and Lidz et al. (2011) conducted experiments in which participants had to verify visual scenes that sometimes had advantages (in speed or accuracy) for possible strategies associated with the meaning of *most*: a strategy Pietroski et al. called *OneToOnePlus*, which requires pairing objects A that have property P with objects A that don't have property P, and a *selection strategy* that requires estimating the cardinalities of sets of all objects that are being compared. They have found that participants did not make use of these strategies when the setup made them advantageous, but rather used the subtraction strategy throughout the experiment. They interpret this result as supporting the Interface Transparency Thesis, which states that "the verification procedures employed in understanding a declarative sentence are biased towards algorithms that directly compute the relations and operations expressed by the semantic representation of that sentence." (Lidz et al., 2011, p. 233).

However, a lot of the questions remain unanswered. While the results of the studies can be interpreted to be indicative of a direct relationship between meaning and verification, the choice of a particular verification strategy could be guided by multiple other factors, such as cognitive load, the presentation of the stimuli (the types of objects in a visual scene, their position on the screen, time of presentation, etc.), a bias towards using the same initially assumed strategy throughout the whole experiment, etc. (see Kotek et al. 2015 for similar arguments).

One such potentially influential factor is working memory load. Previous research on the involvement of working memory in verification of quantified expressions has shown that proportional quantifiers like *more than half* and *most* require higher working memory load than other types of quantifiers (Szymanik and Zajenkowski 2010; Zajenkowski et al. 2011). Furthermore, Steinert-Threlkeld et al. (2015) present data suggesting that verification of *most* and *more than half* may involve working memory to a different extent. The set-up used by Hackl, a self-paced counting task, would be expected to put significant strain on working memory – subjects did not have access to the whole visual scene they had to verify. As Steinert-Threlkeld et al. (2015) point out, the mode of presentation (paired vs. random stimuli, dots vs. letters) impacts the degree of interaction between working memory and verification; for *most* and *more than half*, this impact is different.

Even if the results are taken at face value, it is not clear what this relationship is like without detailed algorithmic-level predictions of how different specifications of truth conditions are computed.

Motivated by these considerations, we will present the results of an experimental study

comparing the verification profiles of *most* and *more than half*, which replicates Hackl's original setup (specifically, his Experiment 1) with some modifications. The results of our experiment suggest that there are some consistent differences in how speakers verify *most* and *more than half* – the former tends to rely more on approximation, and the latter tends to trigger a more precise strategy. However, these differences do not seem to originate from the different specifications of truth condition associated with these quantifiers – the verification patterns we have observed were inconsistent with algorithmic-level predictions based on the specifications of truth conditions alone. Instead, the choice of a particular strategy in our task depended on various factors: both *most* and *more than half* were impacted by individual working memory capacity and the changes in experimental setup from Hackl (2009) that elicited more approximative procedures.

More importantly, our exploratory analysis revealed that three groups of participants can be distinguished based on their preferred strategy. This points to the fact that verification procedures are individualized and flexible – they depend on the type of the task and input, as well as on cognitive resources of subjects. All of these considerations lead us to suggest that instead of triggering a particular default procedure, each quantifier is associated with a collection of verification strategies. Among others, this idea has been previously proposed by Suppes (1982), who argued that the meaning of a sentence can be treated not just as one procedure, but as a collection of those (see also Szymanik 2016). We can expect, then, that some of these procedures overlap for *most* and *more than half*.

2 Current study

2.1 Predictions

Hackl (2009) argues that verifying *most* requires participants to keep track only of the color that is leading at any given moment, the so-called lead-counting strategy. Verifying *more than half*, on the other hand, does not rely on lead-counting: it instead requires keeping track of how many dots the subjects saw in both colors and then comparing the two quantities – using either precise calculations or approximation. This latter procedure is more demanding, as reasoners would have to store a bigger amount of information in their working memory, as well as performing manipulations with it.

Given these considerations, we expect that participants with higher working memory capacity will have shorter reaction times for *more than half* (compared to participants with lower working memory scores), because these participants will be better able to cope with the additional load. At the same time, we expect that subjects with higher memory scores will make fewer mistakes when verifying sentences overall, resulting in a smaller gap in accuracy between *most* and *more than half*[1].

Prediction 1. The higher working memory capacity, the smaller will be the RT effect (the difference in reaction times between most and more than half) and the smaller the accuracy effect (difference in accuracy).

Our second prediction tests Hackl's hypothesis that *most* requires a lead-counting strategy and that a self-paced counting paradigm facilitates such a strategy. On each trial, we coded one screen to have an advantage for the target color: after viewing the first few increments of the scene, which were ambiguous as to what color was leading, subjects saw a screen that

[1] Although overall difference in accuracy was not significant in Hackl's study, subjects were more accurate when verifying *more than half*.

contained a clear advantage for the target color. Assuming that this advantage is irrelevant for verifying *more than half*, we make the following prediction:

Prediction 2. Reaction times on the target screen (namely, screen 4) for *most* will be significantly lower than reaction times on the target screen for *more than half*.

2.2 Participants

Thirty-five (8 female, 24 male, 1 genderfluid) subjects were initially recruited for the study via Prolific.ac, all native speakers of English. Participants were between 18 and 35 years old and were located in the United States. They viewed the experiment in their web browsers, and the average completion time was 14 minutes. Subjects received £2.50 as compensation.

2.3 Materials

The experiment consisted of two sections. In the first section, the digit span task (Schroeder et al., 2012), subjects had to memorize sequences of digits and reproduce them in reverse order. We administered this task as a measure of the working memory capacity of each participant. In the second section, the quantity judgment task, participants had to compare statements such as *Most of the dots are blue* and *More than half of the dots are blue* against visual stimuli, as in Hackl (2009). They were required to press a button for whether the sentence matches the visual stimulus or not. The experiment consisted of 24 target items: 12 sentences with the quantifier *most* and 12 with the quantifier *more than half*. In each group of sentences, 6 of the statements were true (i.e., when the subjects saw the statement *Most of the dots are blue*, it was followed by a scene that matched that description) and 6 were false. The experiment also included thirty-six fillers – sentences with non-proportional quantifiers such as *At most six of dots are yellow*, *Some dots are blue*, *Few dots are green*, etc.

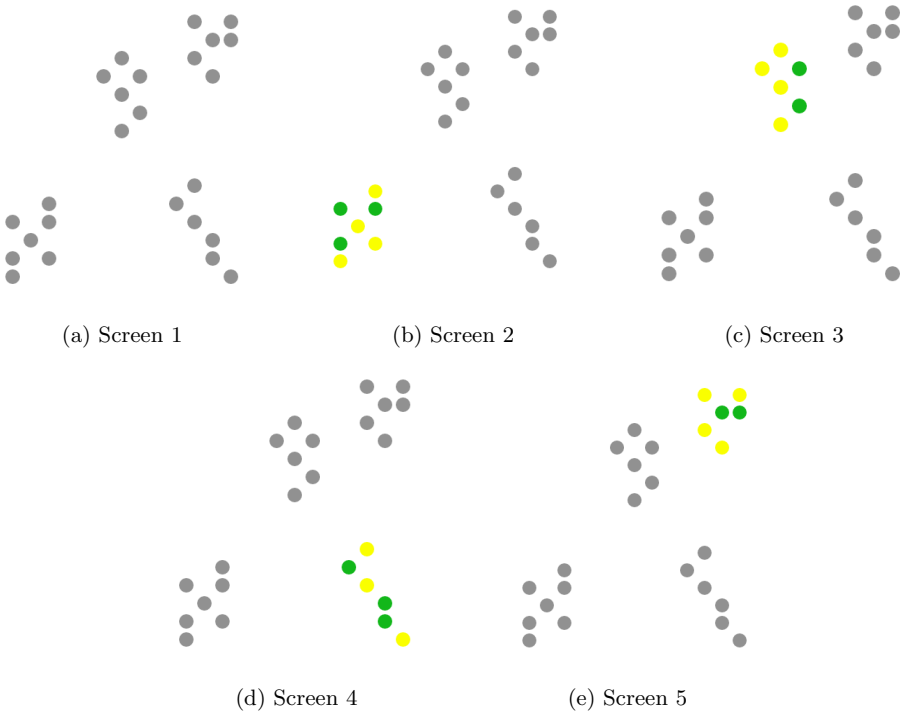
The visual stimuli consisted of pictures of dots scattered across the screen. On the trials with the two quantifiers of interest, it was never clear whether the statement was true or false until screen 5, the last screen. In our analyses below, we focus only on the first 4 screens. On some of the filter trials, it was clear whether the statement was true before reaching the last screen, on the second or third screen. Subjects were informed that they could press one of the answer keys at any point during the trial (see Talmina 2017 for further details).

The main difference between the current setup and Hackl (2009) was the position of the dots – in our experiment they were scattered (see Figure 1), while in Hackl (2009) they appeared in two rows.

2.4 Procedure

For the digit span task, participants saw sequences of digits. Each digit was displayed for 1000 ms. At the end of the sequence participants were given as much time as they needed to fill in the digits in the reversed order. The length of sequences gradually increased and the task ended at the point when participants made three errors in a row. No feedback was given to participants about their performance.

For the sentence judgment task, at the beginning of each trial, subjects saw the sentence that they had to verify in 24pt font on their screen. The time of presentation was not limited, and the subjects had to press the spacebar to proceed to the image. In the first screen (see Figure 1 for an example of a sequence of screens), only the outlines of the dots were visible. The subjects had to press the spacebar to move through the screens, gradually exposing the colors of dots.



(a) Screen 1 (b) Screen 2 (c) Screen 3

(d) Screen 4 (e) Screen 5

Figure 1: Example of a sequence of events in a trial.

When subjects uncovered a new screen, the dots they had previously seen were covered again. They also were not allowed to go back between screens. Participants were informed that they could respond during any part of the trial by pressing the "Y" key (for "yes") on their keyboard if they thought the sentence matched the visual stimuli or the "N" key (for "no") if they thought the sentence was not a correct description of the visual stimuli.

We recorded the information about the time it took the subjects to press a key on every screen, which key was pressed, and whether their response on every trial was correct (but no feedback was given to participants about these).

2.5 Results

Subjects made slightly more mistakes with *most* (in 17.9% of all *most* items) than with *more than half* (in 13.8% of all *more than half* items), but the difference was not significant (Wilcoxon rank-sum test $W = 81576; p = 0.1204$). When analyzing reaction time data, we only looked at the correctly responded trials. Overall reaction times were significantly affected by quantifier: participants took longer verifying *more than half* than *most* when looking at total reaction times ($W = 2043000; p < 0.001$). The latter finding differs from Hackl's, who interpreted the lack of significant overall difference to mean that subjects treated the two quantifiers as equivalent.

Hackl (2009) also reported a difference in reaction times for the first 4 screens (when collapsing across two quantifiers) where participants cannot yet make a decision about whether

the visual scene matches the sentence. In his study, the reaction times gradually increased from screen 1 to screen 4. In our own study, there was also a main effect of screen (Kruskal-Wallis test $H(3) = 140.39; p < 0.001$), but we do not observe such a gradual increase per screen. Instead, in our study participants took significantly longer on screen 1 in comparison to other screens. Focused comparisons of the mean ranks between screens support that: the significant differences were found mostly between screen 1 and other screens.

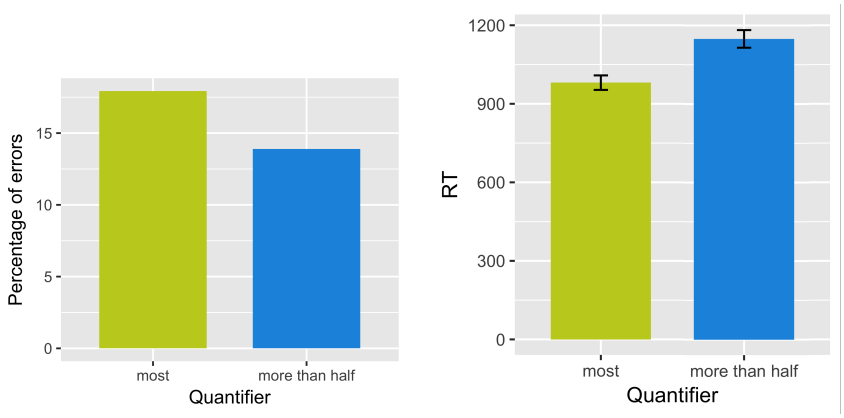


Figure 2: Percentage of errors per quantifier (left plot) and overall reaction times (right plot). The error bars for the reaction times indicate the standard error value.

We further investigated whether there was a screenwise difference in the RT effect – i.e., whether the discrepancies in reaction times between *most* and *more than half* were significantly larger on some screens than on others. To do so, we collected mean reaction times for every subject per every screen, and calculated the difference between *most* and *more than half*. However, no significant differences were found. Thus, the difference between two quantifiers was of similar size for all screens. This goes against Hackl's suggestion that *most* favors lead counting and the prediction that we subsequently made based on this that on the target screen, screen 4, the difference between *most* and*more than half* should be larger than on other screens.

To investigate the relationship between working memory capacity and accuracy and RT effects, we assigned a memory score to every subject based on the length of digit sequences they could remember in the reverse digit span task. We found a negative correlation between memory score and accuracy effect (Pearson's $r = -0.22$[2]), suggesting[3] that a higher memory score is related to lower difference in accuracy between *more than half* and *most*. Similarly, there was some negative correlation between memory score and RT effect (Pearson's $r = -0.2$), which also points to a connection between working memory capacity and lower processing times for more than half, in line with our predictions.

We have noted that the reaction times in the first screen were considerably longer than on other screens (and than RTs in the first screen in Hackl 2009). This might be explained by subjects attempting to estimate the total number of dots (or count them) before proceeding with the task. However, as we can see from Figure 4, not all participants behaved in this way: while some spent over 3000 milliseconds looking at the first screen, others were fast and

[2]We do not provide p-values as they are not considered to be diagnostic for correlation analyses. We only examine the correlation coefficient itself.

[3]See Talmina (2017) for a discussion of why we consider this result to be meaningful.

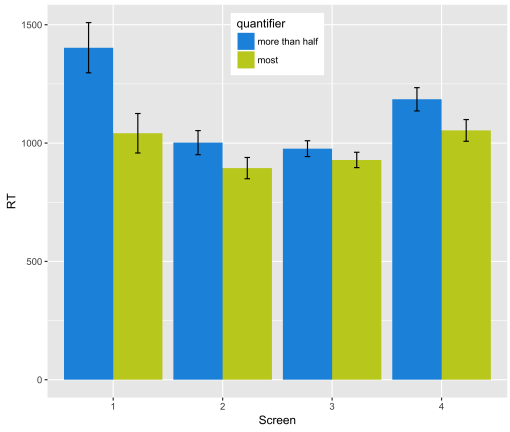


Figure 3: Mean reaction times per screen. The error bars indicate the standard error value.

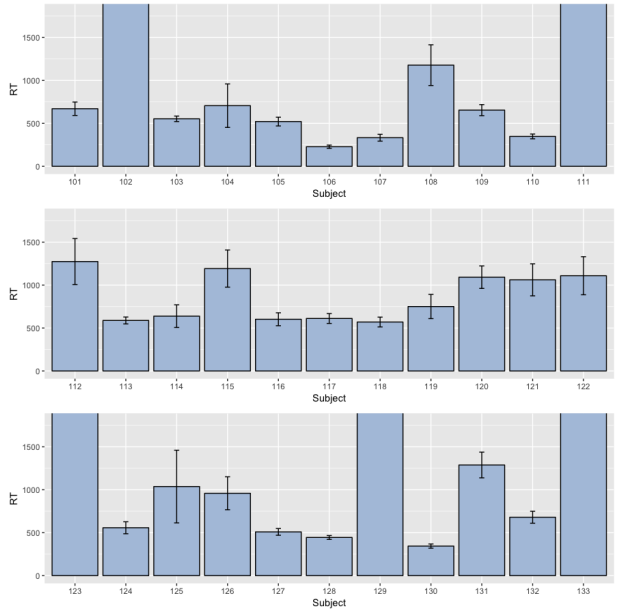


Figure 4: Mean RTs in the first screen for each participant.

took around 500 milliseconds to proceed to the next screen. Another group was in the middle: subjects who spent around 1000 milliseconds in the first screen on average.

As we've argued before, estimating the total number of dots requires additional executive resources, and is most likely justified when participants need to know precisely how many blue dots there need to be for a statement like *More than half of the dots are blue* to be true. In other words, people who look at the first screen for longer are probably going to use a more precise strategy. To explore this intuition, we divided our subjects into three groups based on average

time spent on the first screen: the "counting" group (on average, > 2000 ms spent on the first screen) who we suspect used a precise strategy throughout the whole experiment, the "mixed" group ($1000 - 2000$ ms) who we believe used a mixture of precise and approximative techniques, and the "fast" group (< 1000 ms) who were likelier to use an approximative strategy.

In the "counters" group (5 subjects), we found that there was a significant difference in accuracy between *most* and *more than half* ($W = 2100; p = 0.0148$), but no difference in reaction time ($W = 20757; p = 0.4057$). In the "mixed" group (9 subjects), we found no effect of quantifier on overall accuracy ($W = 4608; p = 1$), but *most* and *more than half* differed significantly in reaction times ($W = 81854; p = 0.012$). In the "fast group" (19 subjects), there was no effect of quantifier on accuracy ($W = 26676; p = 0.4694$), but it was a significant factor for reaction times ($W = 298730; p = 0.00042$).

This analysis was done in a post-hoc manner, so the existence of distinct strategies and their characteristics should be examined in future studies intentionally looking at this aspect.

3 Discussion and conclusion

The results of our experiment diverge in several respects from Hackl (2009). Hackl found no significant overall differences in RTs and accuracy between *most* and *more than half*, which he suggested served as evidence that subjects "treat the two expressions as essentially equivalent" when they are faced with a self-counting task. In the present study, however, we have found that mean difference in overall RTs between the two quantifiers was significant: participants took less time to verify *most* than *more than half*, suggesting that they did not treat in fact the two expressions as equivalent.

We also found a different pattern of screen-by-screen RT change: while Hackl observed a linear decrease in RTs as subjects proceeded through the scene, in our experiment RTs were highest in the first screen, then dropped in screen 2 and remained stable on screen 3, increasing again in screen 4. We built upon this finding to show that participants in our study used distinct verification strategies: while some started moving through the scene right away and relied on more approximative strategies, other spent more time on screen 1, where only the outlines of dots were visible, to count how many there were (see Figure 4).

Further, we investigated Hackl's hypothesis that the verification of *most* requires a lead-counting strategy. We have argued that if such a strategy is employed, subjects would take advantage of additional cues about the leading color that we have supplied in the target screen. As we have found no difference between the presence or absence of the RT effect (the difference in mean RTs between *most* and *more than half*) on the target screen compared to other screens, it is difficult to explain why the presentation mode would help subjects implement a lead-counting strategy, but they would ignore other cues.

The tendency for negative correlation between a subject's memory score and the accuracy effect (the difference between how many errors a subject made when verifying *more than half* and the number of errors she made when verifying *most*) suggests that the higher a subject's working memory capacity was, the fewer mistakes they made when verifying *most*. Higher working memory capacity possibly allowed participants to use a more precise strategy when verifying *most*. Similarly, it allowed reasoners to process *more than half* faster: the algorithm for this quantifier requires storing two numbers in memory, which was easier to do for participants with higher cognitive resources.

In summary, only Prediction 1 (the effect of working memory capacity on accuracy and RT effects) was supported by the data. Prediction 2 was not supported by the data, as we found no significant differences in RTs between *most* and *more than half* on the target screen.

However, we learned other things from our results. Compared to Hackl's results, we do find a difference in overall reaction times between the two quantifiers, which suggests a possible difference in their meanings as well. Moreover, our results highlight potential role of individual differences in competing these tasks (or, perhaps, in verification strategy that is used) which should be investigated in future studies. Hence, we believe that Hackl's experimental conclusions may be premature. There is not enough evidence to claim that *more than half* and *most* are semantically equivalent and associated with different verification strategies. The situation may be much more sophisticated. Each of these quantifiers may be associated with a collection of potential verification strategies which are used in different proportions by different subjects. The choice of the verification strategies depends not only on the quantifier but also on the context, e.g., experimental setup. Crucially, the verification strategies subjects choose seem to be sensitive to general cognitive constraints, like working memory.

References

Hackl, M. (2009). On the grammar and processing of proportional quantifiers: *most* versus *more than half*. *Natural Language Semantics*, 17(1):63–98.

Kotek, H., Sudo, Y., and Hackl, M. (2015). Experimental investigations of ambiguity: the case of most. *Natural Language Semantics*, 23(2):119–156.

Lidz, J., Pietroski, P., Halberda, J., and Hunter, T. (2011). Interface transparency and the psychosemantics of most. *Natural Language Semantics*, 19(3):227–256.

Pietroski, P., Lidz, J., Hunter, T., and Halberda, J. (2009). The meaning of 'most': Semantics, numerosity and psychology. *Mind and Language*, 24(5):554–585.

Schroeder, R. W., Twumasi-Ankrah, P., Baade, L. E., and Marshall, P. S. (2012). Reliable digit span: A systematic review and cross-validation study. *Assessment*, 19(1):21–30.

Steinert-Threlkeld, S., Munneke, G.-J., and Szymanik, J. (2015). Alternative representations in formal semantics: A case study of quantifiers. In *Proceedings of the 20th Amsterdam Colloquium*, pages 368–377.

Suppes, P. (1982). Variable-free semantics with remarks on procedural extensions. *Language, Mind and Brain*, pages 21–34.

Szymanik, J. (2016). *Quantifiers and Cognition: Logical and Computational Perspectives*. Springer.

Szymanik, J. and Zajenkowski, M. (2010). Quantifiers and working memory. In *Logic, Language and Meaning*, pages 456–464. Springer.

Talmina, N. (2017). Quantifiers and verification strategies: connecting the dots (literally). Master's thesis, ILLC, Universiteit van Amsterdam.

Zajenkowski, M., Styła, R., and Szymanik, J. (2011). A computational approach to quantifiers as an explanation for some language impairments in schizophrenia. *Journal of Communication Disorders*, 44(6):595–600.

Asserting a scalar ordering: The non-temporal interpretation of *mae*[*]

Yuta Tatsumi

University of Connecticut, Storrs, Connecticut, USA
yuta.tatsumi@uconn.edu

Abstract

This paper argues that a non-temporal interpretation of *mae* 'before' in Japanese includes scalar ordering, but the scalar aspect of the non-temporal interpretation is neither an implicature nor a presupposition. The non-temporal *mae* 'before' asserts a scalar ordering of two propositions, with respect to an attitude holder's belief about degrees of precision, instead of temporal precedence. It will also be argued that the non-temporal *mae* is syntactically different from the temporal one in that the former takes a VP as its complement.

1 Introduction

As shown in (1), Japanese has the temporal connective *mae* 'before', which determines a temporal order between two events.[1] In what follows, Δ stands for a covert noun coindexed with the subject of the matrix clause.

(1) Temporal use of *mae* 'before'
[Δ_1 *nemuru*] *mae-ni* [*Mary*$_1$*-wa* *tegami-o* *kaita*]
sleeps before-LOC Mary-TOP letter-ACC wrote
'Mary wrote a letter before sleeping.'

This study focuses on a non-temporal use of *mae* 'before', which has received little attention in the literature. (2) is an example of the non-temporal use of *mae* 'before'.

(2) Non-temporal use of *mae* 'before'
[Δ_1 *kyoosi* *dearu*] *mae-ni* [*Mary*$_1$*-wa* *kenkyuusya* *desu* *yo*]
teacher COP before-LOC Mary-TOP researcher COP.POL SFP
Lit. 'Mary is a researcher before she is a teacher.'
'Mary is more a researcher rather than a teacher.'

It has been observed that *mae* only combine with a non-stative predicate under the temporal interpretation. (See [10] for an analysis of this property of the temporal use of *mae*.) Due to this selectional property of *mae*, (2) cannot receive a temporal interpretation, and only the non-temporal interpretation is available.

[*]I would like to thank Stefan Kaufmann, Magdalena Kaufmann, and the audience at TaLK 2016 held at the Institute of Cultural and Linguistic Studies, Keio University for their comments and suggestions. Many thanks also to Pietro Cerrone, Sabine Laszakovits, Gabriel Martínez Vera, Roberto Petrosino for their judgments and discussion of the data.

[1]The abbreviations used in this article are as follows: ACC = accusative; COP = copular; CONJ = conjunctive form; GEN = genitive; LOC = locative marker; NEG = negation; NOM = nominative; POL = politeness marker; PRS = present tense; PST = past tense; SFP = sentence final particle; TOP = topic marker

Under the non-temporal interpretation, we compare two propositions with respect to degrees of precision, instead of temporal precedence. For example, (2) is felicitous in the following context.

(3) Context: Mary is a faculty member of the linguistic department. She has too many classes to teach, and she does not have time to do her own research. One day, John, Mary's friend in the department, made a complaint about her working condition to the head of the department.

In this context, a relevant scale is related to speaker's belief about degrees of precision. Hanako believes that the proposition that Mary is a researcher is more precise than the proposition that Mary is a teacher.

The non-temporal meaning of an item meaning "before" is observed in Romance languages such as Italian and Spanish. (4) is an example from Italian.

(4) Italian
Maria è una ricercatrice prima di essere un insegnante.
Maria is a researcher before of to.be a teacher
'Maria is more a researcher rather than a teacher.'

This paper argues that a core aspect of the non-temporal interpretation of *mae* 'before' is scalar ordering. By comparing the non-temporal use of *mae* with the metalinguistics and the epistemic comparatives, it will be shown that the scalar meaning of the non-temporal interpretation is neither an implicature nor a presupposition. Rather, the non-temporal *mae* 'before' asserts a scalar ordering of two propositions with respect to attitude holder's belief.

2 Characteristics of the non-temporal use of *mae*

2.1 The non-temporal use vs. the temporal use

There are pieces of evidence that the non-temporal use of *mae* must be distinguished from the temporal use. First, the non-temporal *mae* cannot be modified by a measure phrase. As shown in (5a), a measure phrase can precede *mae* under the temporal interpretation. In (5a), the measure phrase specifies the range of a temporal gap between two events.

(5) a. [Δ_1 *nemuru*] *ichi-zikan mae-ni* [*Mary*$_1$*-wa nikki-o kaku*]
sleeps one-hour before-LOC Mary-TOP diary-ACC write
'Mary writes her diary one hour before she sleeps.'

b. *[Δ_1 *kyoosi dearu*] *ichi-zikan mae-ni* [*Mary*$_1$*-wa kenkyuusya dearu*]
teacher COP one-hour before-LOC Mary-TOP rsearcher COP

On the other hand, the non-temporal *mae* cannot co-occur with a measure phrase, as in (5b).

Second, the non-temporal *mae* requires a predicate in its complement clause be a present tense form. Under the temporal interpretation, a past tense form cannot be used when a predicate immediately precedes *mae*, as in (6a). However, if *mae* and its complement clause are not adjacent to each other, this restriction is relaxed, as in (6b).

(6) Temporal use of *mae* 'before'

a. *[*John-ga* *kita*] *mae-ni* *Mary-wa* *kaetta*
John-NOM came before-LOC Mary-TOP went.home
'Mary went home before John came.'

b. [*John-ga* *kita*] ***sono*** *mae-ni* *Mary-wa* *kaetta*
John-NOM came its before-LOC Mary-TOP went.home
Lit. 'Mary went home its before John came.'
'Mary went home before the time when John came.'

In contrast, the non-temporal *mae* cannot co-occur with a past tense form, even when a complement clause is not adjacent to it. As shown in (7b), a past tense form makes the sentence ungrammatical even if there is an intervening item between *mae* and its complement clause.

(7) Non-temporal use of *mae* 'before'

a. *[Δ_1 *kasyu* *deatta*] *mae-ni* *Mary$_1$-wa* *hahaoya* *deatta.*
singer COP.PST before-LOC Mary-TOP mother COP.PST
'Mary was more of a researcher rather than she was a teacher.'

b. *[Δ_1 *kasyu* *deatta*] ***sono*** *mae-ni* *Mary$_1$-wa* *hahaoya* *deatta.*
singer COP.PST its before-LOC Mary-TOP mother COP.PST
Lit. 'Mary was a singer before the time when she was a mother'
'Mary was more of a researcher rather than she was a teacher.'

Third, an external argument cannot be realized in an adverbial clause under the non-temporal use of *mae*. As shown in (8a), an external argument can be realized in a temporal clause. On the other hand, non-temporal *mae* does not allow the presence of an external argument in its complement clause, as in (8b). Notice also that although an external argument cannot be overtly realized in a non-temporal clause, it must be interpreted as the same referent of the external argument in the main clause, as can be seen in the translation of (2).

(8) a. Temporal use of *mae* 'before'
Mary$_1$-wa [*kanozyo$_1$-ga* *sinu*] *mae-ni* *isyo-o* *kaita.*
Mary-TOP she-NOM die before-LOC will-ACC wrote
'Mary wrote her will before she died.'

b. Non-temporal use of *mae* 'before'
Mary$_1$-wa [*kanozyo$_1$-ga* *kyoosi* *dearu*] *mae-ni* *kenkyuusya* *dearu.*
Mary-TOP she-NOM teacher COP before-LOC rsearcher COP
'Mary is more a researcher rather than a teacher.'

These data indicate that the non-temporal interpretation of *mae* is not derived from its temporal interpretation. The non-temporal use must be distinguished from the temporal use.

2.2 The non-temporal use vs. metalinguistic comparatives

One might consider that the non-temporal interpretation is an example of metalinguistic comparatives (for analyses of metalinguistic comparatives, see [1], [9], [2], and references therein). However, there are some properties observed only in the non-temporal use of *mae*.

First, the non-temporal use of *mae* is different from metalinguisitic comparatives (hereafter MCs) in that compared propositions are entailed in the non-temporal use of *mae*. [1] and [9] observe that a compared proposition is not entailed in English and Greek MCs, and hence

cancelable. Japanese MCs also have the same property. (9a) is an example of Japanese MCs. (9b) denies the proposition that Mary is a teacher, and (9c) denies the proposition that Mary is a researcher. These sentences can be uttered after (9a).

(9) a. Metalinguistic comparatives
Mary-wa kyoosi toyuu-yori kenkyuusya dearu.
Mary-TOP teacher COMP.say-than rsearcher COP
'Mary is more of a researcher rather than she is a teacher.'

b. *honntoo-wa kyoosi-de-wa nai kedo ne.*
really teacher-COP-TOP NEG but SFP
'To tell the truth, she is not a teacher, though.'

c. *honntoo-wa kenkyuusya-de-wa nai kedo ne.*
really researcher-COP-TOP NEG but SFP
'To tell the truth, she is not a researcher, though.'

In contrast, compared propositions in the non-temporal *mae* cannot be canceled. (10b,c) are infelicitous after (10a).

(10) a. Non-temporal use of *mae*
[Δ_1 *kyoosi dearu*] *mae-ni* [*Mary*$_1$*-wa kenkyuusya dearu*]
teacher COP before-LOC Mary-TOP rsearcher COP
'Mary is more a researcher rather than a teacher.'

b.# *honntoo-wa kyoosi-de-wa nai kedo ne.*
really teacher-COP-TOP NEG but SFP
'To tell the truth, she is not a teacher, though.'

c.# *honntoo-wa kenkyuusya-de-wa nai kedo ne.*
really researcher-COP-TOP NEG but SFP
'To tell the truth, she is not a teacher, though.'

The infelicity of (10b,c) shows that compared propositions are entailed. There is another piece of evidence that compared propositions must be entailed under the non-temporal interpretation of *mae*. (11) is an example of German MCs, excerpted from [8].

(11) German
Das is eher eine japanische als eine chinesische Maschine.
this is more a Japanese than a Chinese machine
'This is more likely a Japanese than a Chinese machine.'

A similar example of MCs in Japanese is given in (12a). However, we cannot express the same meaning by using the non-temporal *mae*, as can be seen in (12b).

(12) a. Metalinguistic comparatives
kore-wa kokusansya to-yuu yori gaisya dearu.
this-TOP domestic.car COMP-say than foreign.car COP
'This is more a foreign car rather than a domestic car.'

b. Non-temporal use of *mae*
* *kore-wa kokusansya dearu mae-ni gaisya dearu.*
this-TOP domestic.car COMP-say than foreign.car COP
Int. 'This is more a foreign car rather than a domestic car.'

I suggest that (12b) is unacceptable because a single car generally cannot be a domestic one and a foreign one at the same time. Under the present analysis, compared propositions are presupposed in the non-temporal *mae*. If compared propositions cannot be true at the same time, the resulting sentence becomes unacceptable. This restriction is not observed in MCs. The contrast between (12a,b) thus shows that compared propositions in the non-temporal use of *mae* are entailed, in contrast to MCs.

Second, only a nominal predicate can be used in the complement clause of the non-temporal *mae*. [9] observes that Englich MCs are cross-categorial and can compare different syntactic categories. Again, the same behavior holds in Japanese MCs. In (13a), an adjectival predicate is used, and the sentence is acceptable. On the other hand, when an adjectival predicate occurs in the complement clause of the non-temporal *mae*, the resulting sentence is unacceptable, as in (13b). Remember that temporal use of *mae* is incompatible with stative predicates, and the temporal interpretation is unavailable in (13b).

(13) a. Metalinguistic comparatives
[Δ_1 [$_{AP}$ *kasikoi*]] *to-yuu* *yori* [*Mary*$_1$*-wa* [$_{AP}$ *zurui*]]
clever COMP-say than Mary-TOP sly
'Mary is more clever than sly.'

b. Non-temporal use of *mae*
?*[Δ_1 [$_{AP}$ *kasikoi*]] *mae-ni* [*Mary*$_1$*-wa* [$_{AP}$ *zurui*]]
clever before-LOC Mary-TOP sly
Int. 'Mary is more clever than sly.'

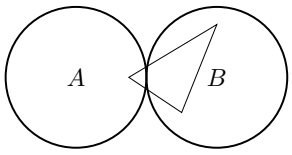
Notice that this restriction on the type of predicate holds in the complement clause of *mae*, but not in the matrix clause that an adverbial clause attaches to. In (14a), an adjectival predicate is used in the complement clause of *mae*. In this case, the sentence is unacceptable even though the matrix predicate is nominal. On the other hand, if a nominal predicate is used in the complement clause of *mae*, the sentence is acceptable as in (14b). Notice that the matrix predicate is an adjective in (14b).

(14) a.?*[Δ_1 [$_{AP}$ *kasikoi*]] *mae-ni* (*somosomo*) [*Mary*$_1$*-wa* [$_{NP}$ *kinben*] *dearu*]
clever before-LOC to.begin.with Mary-TOP industrious COP
Int. '(To begin with,) Mary is more industrious than clever.'

b. [Δ_1 [$_{NP}$ *kinben*] *dearu*] *mae-ni* (*somosomo*) [*Mary*$_1$*-wa* [$_{AP}$ *kasikoi*]]
industrious COP before-LOC to.begin.with Mary-TOP clever
'(To begin with,) Mary is more clever than industrious.'

Verbal predicates also cannot be used in the complement clause of the non-temporal *mae*. Let us consider the figure in (15).

(15)



We can describe (15) by using a MC as in (16a). On the other hand, the non-temporal *mae* cannot be used to describe the figure in (15). (16b) is unacceptable.

(16) a. Metalinguistic comparatives
sankaku-wa en A-ni kasanat-teiru to-yuu yori
triangle-TOP circle A-with overlap-ASP COMP-say than
en B-ni kasanat-teiru.
circle A-with overlap-ASP
'A triangle is overlapping with the circle B rather than the circle A.'

b. Non-temporal use of *mae*
* *sankaku-wa en A-ni kasanat-teiru mae-ni en B-ni kasanat-teiru.*
triangle-TOP circle A-with overlap-ASP before-LOC circle A-with overlap-ASP
'A triangle is overlapping with the circle B rather than the circle A.'

Notice that compared propositions in (16) are independently true, and the semantic requirement of the non-temporal use of *mae* is satisfied. (17) is true in the context.

(17) *sankaku-wa* { *en A-ni* | *en B-ni* } *kasanat-teiru.*
triangle-TOP circle A-with circle B-with overlap-ASP
'A triangle is overlapping {with the circle A | with the circle B}.'

This means that there is no presupposition failure in (16). Nevertheless, (16b) is unacceptable. The unacceptability of (16b) is can be captured if we assume that the non-temporal use of *mae* requires a nominal predicate in its complement clause. It seems that a similar restriction on predicates is observed in Italian. My consultant found that (18) is unacceptable, although the degradedness of (18) seems to be subject to speaker variation.

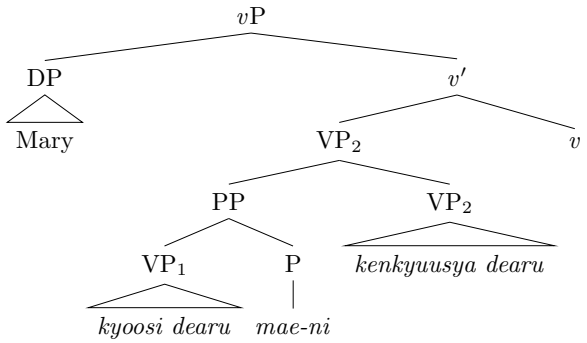
(18) Italian
?* *Giovanni è inelligente prima di essere giovane.*
Giovanni is clever before of to.be young
Int. 'Giovanni is more clever rather than young.'

To sum up, the data in this section indicate that the non-temporal use of *mae* is different from MCs in several respects.

3 Analysis

I propose that the non-temporal *mae* 'before' takes a VP as its complement. A *mae*-clause then attaches to another VP. The structure of (2) is given in (19). Following [3], I assume that an external argument is merged as a specifier of vP.

(19)



Semantically, I would like to pursue an analysis in which the non-temporal *mae* compares two propositions based on attitude holder's belief about degrees of precision. The semantic denotation of the non-temporal *mae* is given in (20). Here, ε is the type of events and t is the type of truth values. The non-temporal *mae* asserts a scalar ordering, and 'p $<_\pi$ q' roughly means that an attitude holder believes that p is more precise than q.

(20) $[\![mae]\!]^{w,c,g} = \lambda \mathrm{P} \in \mathrm{D}_{<\varepsilon,t>} \,.\, \lambda \mathrm{Q} \in \mathrm{D}_{<\varepsilon,t>} \,.\, \lambda \mathrm{e} \in \mathrm{D}_{\varepsilon} : \mathrm{P(e)} \wedge \mathrm{Q(e)} \,.\, [\mathrm{P(e)} <_\pi \mathrm{Q(e)}]$

The proposal can capture the properties of the non-temporal use of *mae*. The complement clause of the non-temporal *mae* is a VP under the present analysis. I assume that when a VP is not c-commanded by the head of TP, a verb has a present tense form in Japanese. With this assumption, a verb cannot be a past tense form under the non-temporal interpretation of *mae* because there is no TP in the complement clause of the non-temporal *mae*. Notice that the restriction on tense morphology is observed in other languages as well. For example, finite clauses cannot be used in a non-temporal adverbial clause in Italian, as shown in (21).

(21) Italian
* *Maria è una ricercatrice prima che sia un insegnante.*
Maria is a researcher before that be.3SG.PRS.SUBJ a teacher
'Mary is more of a researcher rather than she is a teacher.'

(21) is reminiscent of the restriction observed in the non-temporal use of *mae*. The unacceptability of (21) can be captured by assuming that the non-temporal *prima* also selects a defective clause, but not a full finite clause.

The absence of a TP can capture the other properties of the non-temporal use of *mae*. A measure phrase that specifies the range of a temporal gap between two events cannot be used because there is no information about tense due to the absence of a TP. The non-temporal *mae* takes a VP as its complement, and there is no position for an external argument in an adverbial clause. Therefore, an external argument cannot be realized under the non-temporal interpretation.[2] However, an external argument in the complement clause of the non-temporal *mae* must be coindexed with the subject of the matrix clause. Two VPs semantically share their external arguments because of the AGENT function introduced by the v head.

Remember that two compared propositions must be entailed under the non-temporal interpretation. I suggest that compared propositions cannot be canceled because they are presuppositions, in contrast to MCs.
Under the present analysis, the scalar meaning is encoded in the at-issue meaning. There is a piece of evidence for this assumption. As shown in (22), negation can scope over only the scalar aspect of the non-temporal interpretation.

(22) Non-temporal *mae* 'before'
Mary$_1$*-wa* [Δ_1 *kyoosi dearu*] *mae-ni kenkyuusya de-nai.*
Mary-TOP teacher COP before-LOC researcher COP-NEG
'Mary not is more a researcher rather than a teacher.'

(22) does not mean that the speaker believes that "Mary is not a teacher" is more precise than "Mary is a researcher". Under this interpretation, negation takes scope over the proposition expressed by an adverbial clause. Moreover, negation cannot takes scope over the matrix clause,

[2] See [11] for an analysis of reduced clausal metacomparatives in Greek. It seems that his analysis cannot be applied to the non-temporal use of *mae* because an external argument cannot be overtly realized even when the parallelism requirement on ellipsis is respected.

excluding an adverbial clause. (22) does not mean that the speaker believes that "Mary is a teacher" is more precise than "Mary is not a researcher". If the compared two propositions belong to the at-issue meaning, the unavailability of these interpretations is not expected. Based on this, I assume that compared propositions are presupposed. (22) is true only when the speaker believes that it is not true that "Mary is a researcher" is more precise than "Mary is a teacher". Under this interpretation, negation takes scope only over a scaler aspect of the non-temporal interpretation.

4 A loose end: comparison with epistemic comparatives

It is observed that epistemic comparatives are relativized to an attitude holder ([7]). In this respect, the non-temporal use of *mae* is similar to epistemic comparatives. For instance, a scale ordering is defined according to John rather than the speaker in (23).

(23) a. *John-wa* [*Mary*$_1$*-wa* [Δ_1 *kyoosi dearu mae-ni*] *kenkyuusya dearu*] *to*
John-TOP Mary-TOP teacher COP before-LOC researcher COP COMP
omot-teiru
think-ASP
Lit. 'John thinks that Mary is a researcher before she is a teacher.'
'John thinks that Mary is more a researcher rather than a teacher.'

b. *John-niyoruto* [Δ_1 *kyoosi dearu mae-ni* *Mary*$_1$*-wa kenkyuusya dearu*]
John-according.to teacher COP before-LOC Mary-TOP researcher COP
Lit. 'According to John, Mary is a researcher before she is a teacher.'
'According to John, Mary is more a researcher rather than a teacher.'

The sentences in (23) are true regardless of whether or not the speaker believes that the proposition that Mary is a researcher is more precise than the proposition that Mary is a teacher. John's belief is crucial for the felicity of (23a,b).

However, it seems that the non-temporal use of *mae* is not an example of epistemic comparatives. First, [7] observes that the simple indicative present cannot be used in Italian epistemic comparatives. (24) is excerpted from [7].

(24) Italian
* *Gianni è in ufficio piuttosto che a casa.*
Gianni is in office sooner than at home
Int. 'It is more plausible$_{\text{speaker}}$ that Gianni is at work than at home.'

As for the non-temporal use of *prima*, this type of restriction is not observed. An example of the non-temporal use of *prima* is repeated here as (25).

(25) Italian
Maria è una ricercatrice prima di essere un insegnante.
Maria is a researcher before of to.be a teacher
'Maria is more of a researcher rather than she is a teacher.'

Second, [5] observes that German epistemic comparatives are incompatible with first person desire reports. Their example is given in (26).

(26) German
#*Ich will eher nach Wien fahren als in Bregenz bleiben.*
I want EHER to Wien travel than in Bregenz saty
Int. 'I prefer go to Vienna than stay in Bregenz.'

[5] argues that (26) is infelicitous because of a semantic conflict. Epistemic comparatives are felicitous when there is no direct evidence. However, preference generally requires self-awareness, and self-awareness is seen as direct evidence when it comes to a comparison of epistemic confidence in one's desires. Importantly, this kind of conflict is not observed in the non-temporal interpretation of *mae*, as in (27).

(27) Non-temporal use of *mae* 'before'
[Δ_1 *kasyu dearu*] *mae-ni* *watasi*$_1$*-wa zyoyuu deari-tai.*
singer COP before-LOC I-TOP actress COP.CONJ-want
'I want to be more an actress rather than a singer.'

(27) is acceptable and roughly means that the speaker wants to make the proposition "I am an actress" to be a more precise description than the proposition "I am a singer". Notice also that German epistemic comparatives cannot be used in the context given in (3), as in (28).

(28) Context: Mary is a faculty member of the linguistic department. She has too many classes to teach, and she does not have time to do her own research. One day, John, Mary's friend in the department, made a complaint about her working condition to the head of the department.

#*Marie ist eher eine Forscherin als sie Lehrerin ist.*
Marie is more a researcher than she teacher is
'Mary is more a researcher than a teacher.'

Based on these data, I conclude that the non-temporal interpretation discussed in this paper should be distinguished from epistemic comparatives, although they share some properties.

Lastly, let me make a brief comment on an asymmetry between BEFORE and AFTER, regarding availability of the non-temporal uses. To the best of my knowledge, there is no language in which AFTER derives a non-temporal interpretation like BEFORE. It has been observed that there are asymmetries between BEFORE and AFTER (see [4], [6] and references therein). The non-temporal uses may be counted as another asymmetry between BEFORE and AFTER.

References

[1] Giannakidou Anastasia and Melita Stavrou. Metalinguistic comparatives and negation in Greek. In Jeremy Hartman Claire Halpert and David Hill, editors, Proceedings of the 2007 workshop on Greek syntax and semantics, pages 57–74. MIT Press, Cambridge, 2009.

[2] Giannakidou Anastasia and Suwon Yoon. The subjective mode of comparison: Metalinguistic comparatives in Greek and Korean. Natural Language and Linguistic Theory, 29:621–655, 2011.

[3] Kratzer Angelika. Severing the external argument from its verb. In Johan Rooryck and Laurie Zaring, editors, Phrase structure and the lexicon, pages 109–137. Kluwer, Dordrecht, 1996.

[4] Beaver David and Cleo Condoravdi. A uniform analysis of *before* and *after*. In Rob Young and Yuping Zhou, editors, Proceedings of SALT 13, pages 37–54. 2003.

[5] Herburger Elena and Aynat Rubinstein. Is 'more possible' more possible in German? In Mia Wiegand Todd Snider, Sarah D'Antonio, editor, Proceedings of SALT 24, pages 555–576. LSA and CLC Publications, 2014.

[6] del Prete Fabi. A non-uniform semantic analysis of the italian temporal connectives *prima* and *dopo*. Natural Language Semantics, 16:157–203, 2008.

[7] Goncharov Julie and Monica A. Irimia. Modal comparatives: a cross-linguistic picture. handout, GLOW 40, 2017.

[8] von Fintel Kai and Angelika Kratzer. Modal comparatives: two dilletantes in search of an expert. In Luka Crnič and Uli Sauerland, editors, The Art and Craft of Semantics: A Festschrift for Irene Heim, pages 175–179. 2014.

[9] Morzycki Marcin. Metalinguistic comparison in an alternative semantics for imprecision. Natural Language Semantics, 19:39–86, 2011.

[10] Kaufmann Stefan and Misa Miyachi. On the temporal interpretation of Japanese temporal clauses. Journal of East Asian Linguistics, 20:33–76, 2011.

[11] Lechner Winfried. Metacomparatives: Comments on 'Metalinguistic contrast in the grammar of Greek' by Giannakidou & Stavrou. In Jeremy Hartman Claire Halpert and David Hill, editors, Proceedings of the 2007 workshop on Greek syntax and semantics, pages 75–91. MIT Press, Cambridge, 2009.

Expletive-free, concord-free semantics for Russian *ni*-words

Daniel Tiskin

Saint Petersburg State University
daniel.tiskin@gmail.com

Abstract

The paper presents a puzzle about the licensing of the NPI bisyndetic coordinator *ni... ni* in Russian: being in many respects similar to other Russian NPIs with *ni-* as prefix, it does not require negation to be present when it conjoins VPs. Starting with the hypothesis that the semantics of *ni-* remains constant across different uses, I adapt the mechanism of NPI licensing proposed by Chierchia [2] to the needs of the puzzle. The central idea of the proposal is that a token of *ni* splits the composition process into two processes running in parallel until the null operator licensing NPIs unites them back. Negation, whenever present, affects only one of those processes, whose result is then checked by the null operator.

1 Introduction

Lexical items "dependent" [4] on the presence of negation come in various sorts. Some of them, like English *any*, cannot themselves convey the negative meaning. Others, e.g. Russian *nikto* 'nobody' or *nikogda* 'never', can—at least on the surface—be the only manifestation of negativity in an (elliptical) sentence; normally, however, they too are licensed in environments where an overt token of negation is present. Therefore, an interpretation suggests itself where all those items are treated as semantically non-negative.

On the other hand, it turns out that at least one Russian *ni*-item, namely the bisyndetic coordinator *ni... ni*, has uses where it need not, and in fact cannot, be licensed by negation. Such cases suggest that *ni... ni* itself has negative force, but this is hard to reconcile with the strong intuition that the regular negation *ne* should not be denied its own negative force where *ni... ni* and *ne* co-occur. In such cases one has to show how the double negation effect, which is not observed, is avoided. Another question is how *ni... ni* is licensed in the absence of *ne*.

The present paper offers an account that addresses both issues and does not need to postulate expletive negation or purely NC items, giving both *ne* and *ni* non-trivial and constant denotations (although *ni*-words still do not express negation themselves). The structure of the paper is the following: Section 2 presents the Russian data that will be relevant in the discussion that is to follow. Section 3 briefly outlines the view on the positions of raised elements assumed throughout the paper. Section 4 first presents the mechanism of parallel interpretation in two dimensions, which is induced by the presence of *ni*-items, and subsequently explains the semantic contribution of the negation and of the null operator O_D that is assumed to be the licensor of *ni*-words. The relation between *ni*-items and negation will be indirect and explicated in semantic rather than syntactic terms. The section ends with the examination of cases where the negation is absent, contrasting one ungrammatical and one grammatical example. Finally, Section 5 presents a further puzzle regarding the range of contexts where negation has to co-occur with *ni... ni*.

2 N-words in Russian

The morpheme *ni* has at least two major uses in Russian. First, it is a prefix creating "N-words", or negative concord (NC) items that meet the requirements for strong NPIs [4]. This means that they are licensed in the contexts of sentential negation, where their presence does not result in the semantics of double negation (1), and that they can be used in isolation as a negative response to a question (2).

(1) Večerom ja ničego ne el.
evening.ADV I NI.thing.GEN NEG ate
'I ate nothing in the evening.'

(2) A: — Komu ty podaril cvety?
whom you gave flowers
'To whom did you give (the) flowers?'

B: — Nikomu.
nobody.DAT
'(I gave them to) nobody.'

Second, *ni... ni* is a paired coordinator used in NC contexts (3). The distribution of *ni*-words and of DPs conjoined by *ni... ni* is the same in most cases (compare (1) and (3)).

(3) Večerom ja ne el ni supa, ni kartoški.
evening.ADV I NEG ate NI soup.GEN NI potatoes.GEN.
'In the evening, I ate neither soup nor potatoes.'

However—and this observation is, to the best of my knowledge, new—licensing by negation is not required for *ni... ni* when it conjoins VPs (4).[1]

(4) Čto že kasaetsja Pilata, rešenie èto ego ni ogorčilo, ni obradovalo.
what PRT regards P. decision this he.ACC NI disappointed NI pleased
'As for Pilate, this decision neither disappointed nor pleased him.' (RNC)

This observation may be taken as counterexample to Giannakidou's [3] claim that "...Slavic n-words are ungrammatical without negation. This implies that they are unable to contribute negation on their own, as West Germanic n-words do, despite the fact that their morphological make-up seems to have a negative component" (p. 366).[2]

Importantly, the mere fact that a sentence has a VP conjunction with *ni... ni* does not safely license *ni*-items in other parts of the structure.[3]

[1] To avoid confusion, it should be noted that negating each conjunct by means of the regular negation *ne* is also possible in cases like (4):

(i) Rešenie èto ego ne ogorčilo i ne obradovalo.
decision this he.ACC NEG disappointed and NEG pleased.

[2] Note, however, that cases like (4) have been decreasing in frequency over the last centuries. E.g. a query for the Russian National Corpus (RNC) reveals that single-word VPs conjoined by *ni... ni* have the frequency of about 2.8 ipm in 18th century texts, but only 0.8 ipm in 19th, less than 0.3 ipm in 20th and about 0.1 ipm in 21st century texts.

[3] Again, in the 18th century this seems to have been otherwise:

(i) Nikto ni obvinën, ni opravlen; dvor dan orderom.
NI.person NI accused NI acquitted; household given by.order
'No one has been accused or acquitted; the household was given by order' (RNC, 1755–1757)

(5) ??Nikto ni p'ët, ni est.
NI.person NI drinks NI eats.
Intended: 'No one either drinks or eats.'

If it is the negative particle itself, and not a phonologically null head (as suggested by Zeijlstra [9]) that contributes the semantics of negation in the NC-free sentences such as (6), then we may try to extend this analysis to (1) and (3)—or even to (2), assuming some kind of ellipsis.[4] This would require that we do not treat *ni*-words as genuinely negative. This in turn creates a problem for the analysis of (4).

(6) Vasja ne prisël.
V. NEG came.
'Vasya did not come.'

My motivation in what follows will therefore be to develop an account that would associate some sort of genuine negativity both with *ne* and with *ni*-words and take care of the possibility of their co-occurrence, which should not lead to double negation readings. The proposed account will, of course, also have to do justice to the fact that unlicensed *ni... ni* is licit only in a limited selection of environments.

One more proviso before we move on. In the view of (4), it may be suggested that in case *ni* and *ne* occur adjacent to each other, some sort of quasi-haplological merger or deletion applies after syntax to avoid something like **ne ni*. The combination *ne ni* (as opposed to *ni ne*, which can appear in such contexts as *ni ljubit, ni ne ljubit* 'neither loves nor fails to love') is indeed hard to find. However, at least in metalinguistic contexts such as *ne ni odnogo, a 34* 'not "not a single" but as many as 34' or *ne ni s čem, a s pozorom* 'not with nothing but with shame' they are occasionally found.[5]

3 Raising of *ne* and *ni*

According to Abels [1], *ni*-words are not NPIs but rather PPIs, "licensed in the specifier position of a particular projection ⟨...⟩ if that projection hosts negation" (p. 12). As the relative order of a *ni*-word and negation is therefore reversed compared to the traditional view, *ni*-words are taken to have universal quantificational force. However, Abels noted (fn. 13) that "[t]his choice is largely aesthetic". In what follows, I will assume that *ni*-items remain within the scope of the raised negation, not outside it, and have the quantificational force of existentials (or, for *ni... ni*, disjunctive semantics).

As for negation, I assume that it raises at LF from its surface position at the main verb to the position above all *ni*-phrases.

4 Double contribution semantics for *ni*

The key point of the analysis is that certain lexical items can introduce an additional dimension to the interpretational process, that there are other items that operate on both dimensions, and

[4] For a discussion of the problems with the ellipsis account of "fragment answers", especially the fact that ellipsis in a negative clause is somehow licensed by a preceding clause without negation, see [7]. Zeijlstra's [9] null negative operator occupies the position above both the NPI and the ellipsis site, thus avoiding the non-identity problem.

[5] Thanks to Manuel Križ for suggesting this type of argument.

that yet others can only operate on one. The first group includes Russian *ni*-words. The second group is represented by the silent operator O_D that licenses *ni*-items, to be introduced later. Its role will be to bring the two dimensions back together, yielding the truth conditions of the familiar sort. If a certain type of mismatch between the two dimensions obtains at that stage, the interpretation crashes. Finally, negation falls into the third group.

4.1 Interpretation in two dimensions

The two dimensions will be represented using ordered pairs; I call the left member of a pair C_1 and the right member C_2. I use p as a variable over such pairs: $p = \langle C_1(p); C_2(p)\rangle$, where $C_1(p)$ and $C_2(p)$ are the two projections of p.

Simplifying somewhat (see e.g. [6] for a recent account of paired coordinators), I will treat *ni... ni* as a single lexical entry. As many coordinators, *ni... ni* is rather unselective in terms of the semantic type of its arguments, therefore its semantics is given for arguments $A, B \in D_{\langle\sigma,\tau\rangle}$ for arbitrary σ, τ.

(7) $[\![\text{ni } A, \text{ ni } B]\!] = \lambda\alpha_\sigma\Big\langle [\![A]\!](\alpha) \vee [\![B]\!](\alpha);\quad [\![A]\!](\alpha) \vee [\![B]\!](\alpha)\Big\rangle$

Analogously, the prefix *ni-* introduces quantification in two dimensions:

(8) $[\![\text{ničego}]\!] = \lambda P\Big\langle \exists x.x \text{ thing}.P(x);\quad \exists x.x \text{ thing}.P(x)\Big\rangle$

Thus, *ni... ni* and *ni-* are treated on a par. Whenever a λ-abstractor binds into the angle brackets, Functional Application [5] applies to both dimensions. However, in some cases my analysis predicts that a pair should interact with a pair, as a sentence may contain more than one *ni*-item:

(9) Ni Vasja, ni Petja ničego ne skazali.
NI V. NI P. NI.thing.GEN NEG said
'Neither Vasya nor Petya said anything.'

Given this, we need a composition rule analogous to FA but suited for the case where one of the two merged constituents is a pair.

(10) a. *FA (unpaired × unpaired)*
If $A \in D_{\langle\sigma,\tau\rangle}$ and $B \in D_\sigma$, then $A(B) \in D_\tau$
b. *FA (paired × unpaired)*
If $p = \langle C_1(p); C_2(p)\rangle$, $C_i(p) \in D_{\langle\sigma,\tau\rangle}$ and $B \in D_\sigma$,
then $p(B) = \langle C_1(p)(B); C_2(p)(B)\rangle$ and $C_i(p)(B) \in D_\tau$
c. *FA (unpaired × paired)*
If $A \in D_{\langle\sigma,\tau\rangle}$ and $p = \langle C_1(p); C_2(p)\rangle$, $C_i(p) \in D_\sigma$,
then $A(p) = \langle A(C_1(p)); A(C_2(p))\rangle$ and $A(C_i(p)) \in D_\tau$

Given the definition of FA and the fact that the negation has raised over the position(s) occupied by DPs after QR, the interpretation of (9) starts as in Figure 1.

As can be seen, *ni*-phrases do not take scope w.r.t. negation and only trivially scopally interact with each other (as $\exists xPx \vee \exists xQx \Leftrightarrow \exists x(Px \vee Qx)$). Moreover, none of them introduces negative force into interpretation. This will play a role in the explanation of why no double negation arises.[6] However, at this point the semantics of negation has not yet been introduced,

[6] The same result is elegantly achieved within the approaches that take NC items to undergo obligatory QR

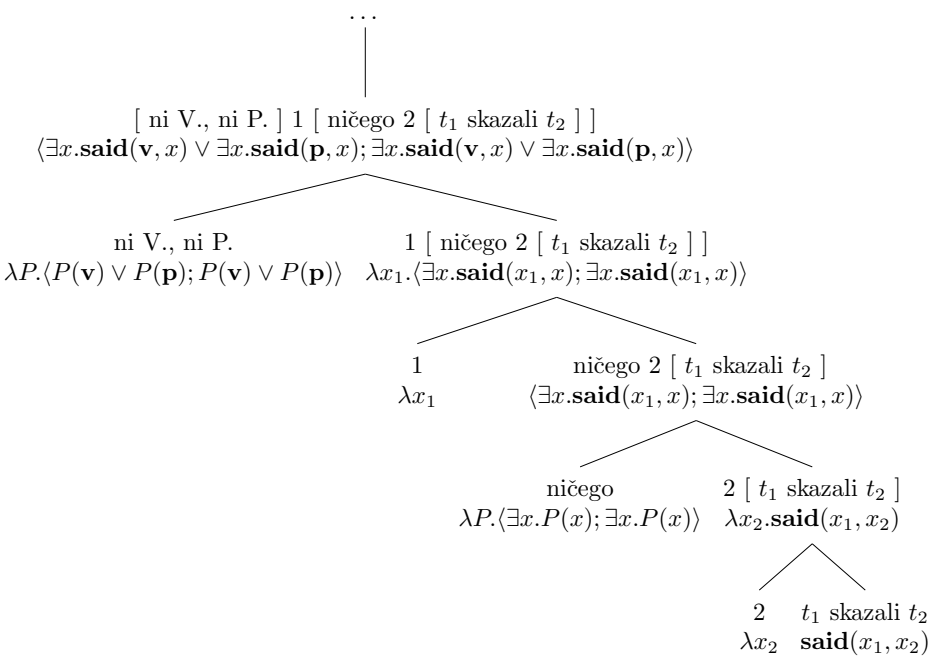


Figure 1: A partial derivation of (9)

and the roles of the two dimensions are unclear, as they do not differ from each other. As will be clarified presently, the role of the second dimension is to "memorize" the denotation of the negation's complement and to transmit it to a later stage of computation.

4.2 Negation and *ni*-licensing

Out of the two semantic dimensions, the negative particle *ne* affects only C_1 of its argument, leaving C_2 intact:

(11) $[\![\neg p]\!] = \langle \neg [\![C_1(p)]\!]; [\![C_2(p)]\!]\rangle$.

Therefore, the step immediately following those shown in Figure 1 will be

(9)′ ne [[ni V., ni P.] 1 [ničego 2 [t_1 skazali t_2]]]
$\langle \neg\big(\exists x.\mathbf{said}(\mathbf{v},x) \vee \exists x.\mathbf{said}(\mathbf{p},x)\big); \exists x.\mathbf{said}(\mathbf{v},x) \vee \exists x.\mathbf{said}(\mathbf{p},x)\rangle$

Following the treatment of *any*-series in Chierchia [2] and Xiang [8], I postulate a silent operator O_D at the left periphery, which licenses *ni*-items. Its semantic contribution invokes the *subdomain alternatives* of its complement's C_1.

Definition 1 (subdomain alternatives)**.** *Given a sentence S whose quantifiers $Q_1, \ldots, Q_n$ range over the domain D, the set of S's subdomain alternatives $\mathcal{ALT}(S)$ is defined as*

$$\mathcal{ALT}(S) = \{S' \mid S' \rightarrow S \text{ and } \exists D' \subset D : Q_1, \ldots, Q_n \text{ range over } D' \text{ in } S'\}.$$

over negation, thus treating them as PPI, such as that of Abels [1].

More specifically (and somewhat differently from Chierchia's account of English), I take it that O_D (a) completes the task left incomplete by negation, i.e. negates its complement's C_2, and (b) asserts all the subdomain alternatives of its complement's C_1 not entailed by C_1. The results of (a) and (b) are conjoined, so the two dimensions collapse back into one.

(12) $[\![O_D\, p]\!] = \neg C_2(p) \wedge \forall S \in \mathcal{ALT}(C_1(p)) : ((C_1(p) \not\to S) \to S$

Intuitively, this performs a check on the relation between the two dimensions: if C_1 as input to O_D does not in some salient sense, cast in terms of entailment, match the negation of C_2, the resulting interpretation will be contradictory (due to conjunction) and the sentence will be ruled out.

Continuing with (9), at the point where O_D is merged we get

(9)″ O_D [ne [[ni V., ni P.] 1 [ničego 2 [t_1 skazali t_2]]]]
$\neg\big(\exists x.\mathbf{said}(\mathbf{v}, x) \vee \exists x.\mathbf{said}(\mathbf{p}, x)\big) \wedge$
$\wedge \forall S \in \mathcal{ALT}\big(\neg\big(\exists x.\mathbf{said}(\mathbf{v}, x) \vee \exists x.\mathbf{said}(\mathbf{p}, x)\big)\big) :$
$\big(\neg\big(\exists x.\mathbf{said}(\mathbf{v}, x) \vee \exists x.\mathbf{said}(\mathbf{p}, x)\big) \not\to S\big) \to S$

Obviously, if neither Vasya nor Petya said anything (where *anything* quantifies over a given domain D), then for any $D' \subset D$ it also holds that neither Vasya nor Petya said anything. Thus all subdomain alternatives of $C_1(p)$, which happens to be 'neither Vasya nor Petya said anything', are entailed by $C_1(p)$. Thus, the whole second conjunct in the denotation of (9)″ is vacuous, and the resulting truth conditions are $\neg\big(\exists x.\mathbf{said}(\mathbf{v}, x) \vee \exists x.\mathbf{said}(\mathbf{p}, x)\big)$.

4.3 The absence of *ne*

Ungrammatical cases. Let us now consider what happens if a sentence does not contain *ne*. We know from Section 2 that in such cases it makes a difference whether *ni* conjoins VPs or occurs elsewhere in the sentence. Let us first consider (13), the ungrammatical pair for (9).

(13) *Ni Vasja, ni Petja ničego skazali.
NI V. NI P. NI.thing.GEN said

We assume that a structure with *ni*-words but without the licensing O_D is syntactically ill-formed (e.g. because O_D has to check the [+ D] feature of *ni*-items, as in [2]). Therefore we grant that (13) does contain O_D. The following predicts that it will be still ruled out due to the lack of *ne*.

Having gone through the steps in Figure 1, we do not add *ne* but proceed directly to

(14) *O_D [[ni V., ni P.] 1 [ničego 2 [t_1 skazali t_2]]]
$\neg\big(\exists x.\mathbf{said}(\mathbf{v}, x) \vee \exists x.\mathbf{said}(\mathbf{p}, x)\big) \wedge$
$\wedge \forall S \in \mathcal{ALT}\big(\exists x.\mathbf{said}(\mathbf{v}, x) \vee \exists x.\mathbf{said}(\mathbf{p}, x)\big) :$
$\big(\exists x.\mathbf{said}(\mathbf{v}, x) \vee \exists x.\mathbf{said}(\mathbf{p}, x) \not\to S\big) \to S$

This denotation minimally differs from that of (9)″ in that there is no negation in C_1. Therefore, none of the elements of the set returned by $\mathcal{ALT}$ is entailed by $\exists x.\mathbf{said}(\mathbf{v}, x) \vee \exists x.\mathbf{said}(\mathbf{p}, x)$: if $\exists x.\mathbf{said}(\mathbf{v}, x) \vee \exists x.\mathbf{said}(\mathbf{p}, x)$ is made true by a single value of x, no proper subdomain of D may be *a priori* assumed to contain that object even if D itself does. So (14) is equivalent to

$$\neg\big(\exists x.\mathbf{said}(\mathbf{v}, x) \vee \exists x.\mathbf{said}(\mathbf{p}, x)\big) \wedge \bigwedge_{S \in \mathcal{ALT}\left(\exists x.\mathbf{said}(\mathbf{v}, x) \vee \exists x.\mathbf{said}(\mathbf{p}, x)\right)} S,$$

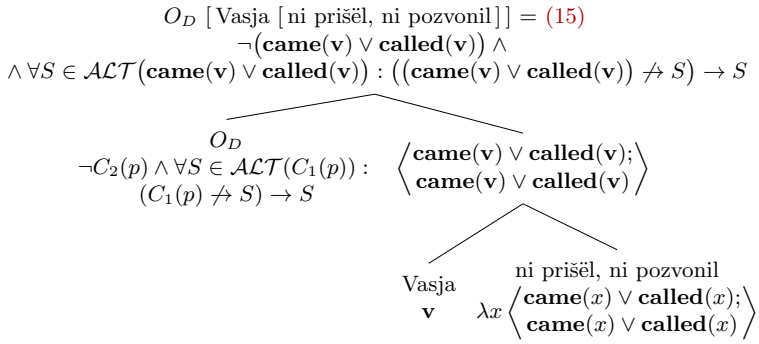


Figure 2: A derivation without *ne*, (15)

which is a contradiction ('in D, neither Vasya nor Petya said anything, but in all $D' \subset D$ one of them did').

To conclude, (14) is correctly predicted to be impossible without *ne* for the same reason as the English **John met any girl* is on Chierchia's approach: applying O_D will yield the contradiction 'for the domain D, John met a girl, but for all $D' \subset D$, he didn't'.

Grammatical cases. Instead of a fairly complicated (4), consider the simpler (15).

(15) Vasja ni prišël, ni pozvonil.
V. NI came NI called
'Vasya neither came nor called.'

Figure 2 shows how (15) is interpreted. The crucial point is that none of the subdomain alternatives of C_1 is entailed by $\neg C_2$ and therefore all the alternatives are negated. This is exactly what we need, since this negation of everything in $\mathcal{ALT}(C_1(p))$ is itself entailed by $\neg C_2$ to the effect that the overall denotation of (15)—the conjunction of that negation and $\neg C_2$—is equivalent to C_2 alone. The quantification over subdomains is again vacuous, since the formula $\mathbf{came}(\mathbf{v}) \vee \mathbf{called}(\mathbf{v})$ contains no quantifiers and therefore entails all its subdomain alternatives (provided that Vasya exists in the given subdomains).

The effect of the proposal is that the absence of *ne* does not prevent (15) from having "negative" meaning, in the same way as the presence of both *ne* and *ni* in (3) does not prevent it from negating, not asserting, that I ate soup or potatoes.

5 Further issues

The puzzle as I presented it in Section 2 was actually simplified. The real distribution of the negation *ne* in the presence of *ni... ni* is somewhat more complex. More precisely, *ne* is not required where in each of the conjoined VPs the verb linearly precedes other material; if some other part of the VP is scrambled in a position before the verb, the presence of *ne* adjacent to the verb is again required:[7]

[7]The situation seems to have been different several centuries ago:

(16) Vy, Stëpka, ni [$_{VP}$ konja ne najdëte], ni [$_{VP}$ nas ne otyščete].
you S. NI horse NEG will.find NI us NEG will.find
'You, Styopka, will neither find a/the horse nor find us.' (RNC)

If only one of the conjoined VPs has preverbal material, *ne* appears only in that conjunct:

(17) mama i brat celyj den' na rabote, a ona ni [kušat' ne prigotovit], ni [uberët].
Mom and brother whole day at work but she NI eat.INF NEG will.prepare NI will.clean.up
'My mom and brother are at work all day long, but she [the wife] would neither make food nor clean up' (RNC)

Similarly, if *ni... ni* conjoins clauses, the main verb, which is then not clause-initial, should be marked with *ne*:

(18) Ni babka moja ni slova ne znala po-ukrainski, ni otec ne znaet.
NI grandmother my NI word NEG knew in.Ukrainian NI father NEG knows
'Neither did my grandmother know a single Ukrainian word nor does my father' (RNC)[8]

As for now, I will have to leave this further puzzle as a direction for future work.

References

[1] Klaus Abels. Expletive (?) negation. In *Proceedings of FASL*, volume 10, pages 1–20, 2002.

[2] Gennaro Chierchia. *Logic in grammar: Polarity, free choice, and intervention*. OUP, 2013.

[3] Anastasia Giannakidou. N-words and negative concord. In M. Everaert and H. van Riemsdijk, editors, *Blackwell Companion to Syntax*, volume III, pages 327–391. Blackwell, 2006.

[4] Anastasia Giannakidou and Hedde Zeijlstra. The landscape of negative dependencies: negative concord and n-words. In Martin Everaert and Henk van Riemsdijk, editors, *The Wiley Blackwell Companion to Syntax*. Blackwell, 2nd ed. edition, 2017.

[5] Irene Heim and Angelika Kratzer. *Semantics in Generative Grammar*. Oxford: Blackwell, 1998.

[6] Moreno Mitrović and Uli Sauerland. Decomposing coordination. In *Proceedings of NELS*, volume 44, pages 39–52, 2014.

[7] Akira Watanabe. The genesis of negative concord: Syntax and morphology of negative doubling. *Linguistic Inquiry*, 35(4):559–612, 2004.

[8] Yimei Xiang. *ONLY*: An NPI-licenser and NPI-unlicenser. *Journal of Semantics*, 34(3):447–481, 2017.

[9] Hedde Zeijlstra. *Sentential Negation and Negative Concord*. PhD thesis, University of Amsterdam, 2004.

(i) Vot i prilëg on na otcovoj mogilke, ni šknët, ni čixnët, ni [$_{VP}$ uxom povedët].
NI will.utter NI will.sneeze NI ear.INS.SG will.move
'So he lay down on his father's grave, making no sound, not sneezing, not moving his ears' (RNC, 1825–1833)

[8] Phrases such as *ni slova* 'not a single word' do, formally speaking, contain *ni* but remain outside the scope of the present inquiry.

The anti-rogativity of non-veridical preferential predicates*

Wataru Uegaki[1] and Yasutada Sudo[2]

[1] Leiden Unviersity
w.uegaki@hum.leidenuniv.nl
[2] University College London
y.sudo@ucl.ac.uk

Abstract

Clause-embedding predicates come in three major varieties: (i) *responsive predicates* (e.g. *know*) are compatible with both declarative and interrogative complements, (ii) *rogative predicates* (e.g. *wonder*) are only compatible with interrogative complements, and (iii) *anti-rogative predicates* (e.g. *hope*) are only compatible with declarative complements. It has recently been suggested that these selectional properties are at least partly semantic in nature. In particular, it is proposed that the anti-rogativity of neg-raising predicates like *believe* comes from the triviality in meaning that would arise with interrogative complements. This paper puts forward a similar analysis for non-veridical preferential predicates such as *hope*. In so doing we also aim at explaining the fact that their veridical counterparts such as *be happy* are responsive.

1 Introduction

Clause-embedding predicates can be classified into three types ([14, 22, 31]):

- Responsive predicates can embed both declarative and interrogative complements, e.g. *know*.
- Rogative predicates can only embed interrogative complements, e.g. *wonder*.
- Anti-rogative predicates can only embed declarative complements, e.g. *believe*.

The main question we would like to tackle in this paper is how this variation should be accounted for. One possibility is to assume that each clause-embedding predicate comes with a lexical specification as to what type of clause it syntactically selects for. A purely syntactic theory of this kind, however, would be unsatisfactory given the stability and predictability of selectional patterns both intra- and cross-linguistically in the sense that predicates that have similar meanings generally exhibit the same selectional properties. For instance, such a theory would not necessarily rule out a version of *know* that is rogative or a version of *wonder* that is responsive. These considerations are taken as evidence that the core selectional properties come from the lexical semantics of the predicates, although idiosyncratic syntactic properties are not necessarily excluded (see [14, 25, 24, 35]).

Recent developments in the area of question semantics have prompted linguists to tackle the issue of complement clause selection from a more semantic perspective, but there are some open issues. In order to illustrate the issue, let us first consider the following type-theoretic approach. One of the standard views of question semantics championed by [19] and others holds that declarative and interrogative clauses denote different kinds of semantic objects.

*We would like to thank Dominique Blok, Kajsa Djärv, Patrick Elliot, Jane Grimshaw, Rick Nouwen, Maribel Romero, Henriette de Swart, and Aaron Steven White for helpful comments and discussion. All remaining errors are our own.

Specifically, declarative clauses denote propositions, while interrogative clauses denote sets of propositions. In this setting, anti-rogative predicates like *believe* can then be analyzed as those whose denotations exclusively select for propositions, and rogative predicates like *wonder* as those whose denotations exclusively select for sets of propositions, as illustrated in (1). Throughout this paper, we write $\hat{\tau}$ for the type of sets of type-τ objects.[1]

(1) a. $[\![\textsf{believe}]\!]^w = \lambda p_{\langle s,t\rangle}.\lambda x_e.\ B_w(x,p)$ b. $[\![\textsf{wonder}]\!]^w = \lambda Q_{\widehat{\langle s,t\rangle}}.\lambda x_e.\ W_w(Q)$

This analysis needs to make an extra assumption about responsive predicates, which are compatible with both types of embedded clauses. The most popular take on this is to assume that when they combine with an interrogative clause, the meaning of the interrogative clause is converted to a specific proposition that represents an 'answer' to the question ([16, 10, 5, 29]). We will put aside the interesting but complicated issue of what counts as an appropriate answer to a question here, but if any such mechanism that converts sets of propositions to propositions is available, it becomes unclear why anti-rogative predicates cannot combine with interrogative clauses. That is, just like (2-a) means roughly 'John doesn't know the true answer to the question *Who danced?*', (2-b) should be able to mean something like 'John doesn't believe the true answer to the question *Who danced?*'.

(2) a. John doesn't know who danced. b. *John doesn't believe who danced.

Some recent theories of question semantics do not make a type distinction between declarative and interrogative clauses ([8, 32, 30]). Specifically, on these accounts, both declarative and interrogative clauses denote sets of propositions, which are taken to represent *issues*, and the difference between declarative and interrogative clauses boils down to whether the issue has been resolved. Then, rogative predicates can be analyzed as those that exclusively select for unresolved issues, anti-rogative predicates as those that exclusively select for resolved issues, and responsive predicates as those that are insensitive to resolvedness. To be more concrete, these restrictions could be encoded as sortal presuppositions as in (3).

(3) a. $[\![\textsf{believe}]\!]^w = \lambda Q_{\widehat{\langle s,t\rangle}}\colon \texttt{resolved}(Q).\lambda x_e.\ B_w(x,Q)$
b. $[\![\textsf{wonder}]\!]^w = \lambda Q_{\widehat{\langle s,t\rangle}}\colon \neg\texttt{resolved}(Q).\lambda x_e.\ W_w(Q)$
c. $[\![\textsf{know}]\!]^w = \lambda Q_{\widehat{\langle s,t\rangle}}.\lambda x_e.\ K_w(Q)$

For such a theory to be truly explanatory, however, it needs to be able to predict which predicates have what restrictions, but this turns out to be a rather vexing issue. For instance, *know* and *believe* have very similar meanings, but why is it that the former is responsive while the latter is anti-rogative?

Recently, [31] and [23] propose a partial answer to this question that concerns the anti-rogativity of neg-raising predicates. As originally noticed by [36], neg-raising predicates are all anti-rogative, e.g. *believe, think, expect, assume, presume, reckon, advisable, desirable, likely.* To explain this robust generalization, [31] and [23] put forward semantic accounts, according to which, such predicates give rise to logically trivial interpretations with interrogative complements, due to their neg-raising property. We do not go into the details of these accounts here, but in our view they are conceptually attractive, as they reduce the selectional properties of these predicates to their independent semantic property. Of course, it needs to be explained which predicates are neg-raising and which ones are not, but that is an independent problem

[1] Since the domain of partial functions of type $\langle\sigma,t\rangle$ and the domain of sets of objects of type-σ are not isomorphic, we will explicitly distinguish sets and their characteristic functions in the present paper.

	Representational	Non-representational (preferential)
Veridical	*know, forget, remember*	*be glad, be surprised, be happy*
Non-veridical	*believe, be certain, doubt*	*hope, wish, demand*

Table 1: Examples of four classes of attitude predicates generated by veridicality and representationality.

that every semantic theory needs to account for.

One limitation of these accounts, however, is that they only explain a subset of anti-rogative predicates. That is, while neg-raising predicates are all anti-rogative, not all anti-rogative predicates are neg-raising. Concretely, predicates like *wish, fear, deny* and *regret* are not neg-raising but are still anti-rogative. It is also interesting to note that English *hope* and its Dutch cognate *hopen* have similar meanings but crucially differ in the neg-raising property ([18]). Nonetheless, both of them are anti-rogative.

In sum, it is conceptually appealing to explain anti-rogativity in semantic terms, and some recent accounts achieve this for neg-raising predicates like *believe*. However, the explanations they offer are not applicable to all anti-rogative predicates. This of course does not mean that these accounts should be dismissed. In fact, we think it is not unlikely that different anti-rogative predicates are anti-rogative for different semantic reasons. In this paper, we will develop a semantic analysis of the anti-rogativity of preferential predicates like *hope* and *fear*. If successful, it will complement the analyses of the anti-rogativity of neg-raising predicates, although we admit that there still will be some anti-rogative predicates that are left unexplained by either account, e.g. *regret* and *deny*.

The idea we will pursue in this paper is similar in nature to the aforementioned accounts of neg-raising predicates: non-veridical preferential predicates like *hope* are prohibited from combining with interrogative clauses, because such combinations are bound to result in trivial meanings. We will furthermore show that this analysis also accounts for the fact that their veridical counterparts like *be happy* are responsive.

2 Veridicality and Anti-Rogativity

Following previous studies on the typology of attitude predicates, especially [1] (see also [6, 15, 34]), we recognize two major semantic classes among them. We say an attitude predicate is *representational* if it expresses a 'propositionally consistent attitudinal state' ([1], p. 3) and is *non-representational* otherwise.[2] For instance, predicates of (un)acceptance (e.g. *know, believe, be certain, deny*) are all representational. Of particular interest for us in this paper are *preferential predicates* which constitute a sub-class of non-representational predicates. Preferential predicates express comparisons of alternatives based on preference orders. They include desideratives (e.g. *hope, wish, want, fear, be surprised, be happy*) and directives (e.g. *demand*).

We observe that veridicality correlates with anti-rogativity in the domain of preferential predicates. We say that a clause embedding predicate V is *veridical* if $\ulcorner \alpha\ V$s that $p \urcorner$ entails $\ulcorner p \urcorner$. Veridicality crosscuts the representational vs. non-representational distinction, giving rise to four classes of attitude predicates, as in Table 1.

[2] [1] lists some linguistic phenomena (namely, mood selection, parenthetical uses, compatibility with epistemic modals) that could be used as diagnostics for representationality of attitude predicates, but space prevents us from discussing them in detail here. It should nonetheless be mentioned that the empirical landscape is not as clear as one might wish. For example, as [1] discusses, mood selection might not be an entirely reliable test, especially given its cross-linguistic variability.

We submit that all non-veridical preferential predicates are anti-rogative. Let us consider some examples. Firstly, non-factive preferential predicates are generally anti-rogative.

(4) a. *Alice prefers which students will be invited to the party.
b. *Ben hopes/wishes which students will be invited to the party.
c. *Chris expects/fears how many students will be invited to the party.

All of these predicates are compatible with finite declarative complements.

(5) a. Alice prefers that Andrew will be invited to the party.
b. Ben hopes/wishes that Becky is invited to the party.
c. Chris expects/fears that Cathy is invited to the party.

It should be remarked that there are preferential predicates that cannot (easily) take finite complements, declarative or interrogative, e.g. such as *want* and *be uneasy*. These predicates are neither anti-rogative nor responsive, and their selectional restrictions need to be somehow lexically stipulated, perhaps as a syntactic condition.

Finally, let us look at preferential predicates that are compatible with interrogative complements. The ones in the following examples are all responsive and veridical (factive, in fact) when combined with declarative complements.[3]

(6) a. Andy is surprised (at/by) which students are invited to the party.
b. Ben is glad/happy which students are invited to the party.
c. Chris liked/hated how many students were invited to the party.

These data corroborate our generalization, but two limitations need to be mentioned. Firstly, the generalization is not exception-less. One notable exception is *regret*, which is factive but anti-rogative. Secondly, there are still some anti-rogative predicates that are not amenable to our generalization or [36]'s generalization concerning neg-raising predicates. For instance, *deny* is neither preferential or neg-raising but is still anti-rogative. These predicates require a yet another account.

Before leaving this section, it should be stressed that our generalization has nothing to say about the veridicality of representational predicates and their selectional properties. However, it is noticeable that veridicality generally implies compatibility with interrogative clauses in the domain of representational predicates as well. As mentioned above, neg-raising matters for non-veridical representational predicates. For instance, *believe* and *predict* are both non-veridical, but the former is anti-rogative, while the latter is responsive. Veridical representational predicates (e.g. *know*), on the other hand, are both non-neg-raising and responsive. We have nothing new to add here, and refer the interested reader to [11], [31] and [23].

3 Why Veridicality Matters for Preferential Predicates

We will now explain why veridical preferential predicates are responsive, while non-veridical ones are anti-rogative. The explanation will be based on the idea that non-veridical preferential predicates with an interrogative clause give rise to trivial meaning while veridical preferentials

[3] Some of these cases sound better with a preposition like *about*, but we should be careful as *about* itself might make an interrogative complement available. For instance, while *think* is anti-rogative, *think about* is responsive. This might or might not be because of the meaning of *about* (see [26] for relevant discussion) or because *think* is neg-raising while *think about* is not. We will leave this issue open here, and avoid examples containing *about*.

don't, regardless of the complement clause-type, assuming (i) a uniform approach to clause-embedding [8, 32, 30] (§3.1) and (ii) the degree-based semantics for preferentials [27] (§3.2).

3.1 A Uniform Approach to Clausal Embedding

We follow [8, 32, 30] and take a uniform approach to clause-embedding where both declarative and interrogative complements denote sets of propositions and all clause-embedding predicates take sets of propositions as arguments. We assume that declarative sentences denote singleton sets of propositions, while interrogative clauses denote non-singleton sets of propositions.[4,5]

(7) a. $[\![\text{Alice jumped}]\!]^w = \{\, \lambda w.\ J_w(a) \,\}$
b. $[\![\text{whether Alice jumped}]\!]^w = \{\, \lambda w.\ J_w(a),\ \lambda w.\ \neg J_w(a) \,\}$
c. $[\![\text{who jumped}]\!]^w = \{\, \lambda w.\ J_w(x) \mid x \in D \,\} \cup \{\, \lambda w.\ \neg\exists x J_w(x) \,\}$

In this setting, representational predicates like *be certain* and *know* have an existential semantics, as in (8).[6]

(8) a. $[\![\text{be certain}]\!]^w = \lambda Q_{\widehat{\langle s,t\rangle}}.\lambda x_e.\ \exists p \in Q[B_w(x,p)]$
b. $[\![\text{know}]\!]^w = \lambda Q_{\widehat{\langle s,t\rangle}}.\lambda x_e\colon\ \exists p \in Q[p(w)].\ \exists p \in Q[p(w) \wedge K_w(x,p)]$

All clause-embedding predicates take a set of propositions, so they are all type-compatible with both interrogative and declarative complements. For instance, the denotation of *believe* is of the same type as that of *know*:

(9) $[\![\text{believe}]\!]^w = \lambda Q_{\widehat{\langle s,t\rangle}}.\lambda x_e.\ \exists p \in Q[B_w(x,p)]$

The anti-rogativity of *believe*, therefore, needs to be explained by other means than type incompatibility. As mentioned before, [31] and [23] propose to reduce it to its neg-raising property. For reasons of space we will not review these analyses here.

3.2 Degree-Based Semantics for Preferential Predicates

Now we are in a position to discuss our analysis of preferential predicates. We follow [27]'s degree-based semantics, which is an adaptation of [34]'s ordering-based analysis of preferentials. The degree-based semantics offers an attractive account of the anti-rogativity of non-veridical preferential predicates with a reasonable assumption about the semantics of degree constructions in general.

Before jumping to the concrete analysis, we mention an important aspect of the semantics of preferential predicates: focus-sensitivity. [34] observes that focus has truth-conditional effects with bouletic predicates ([27]). Here's an example illustrating this (modeled after [27]):

(10) CONTEXT: Natasha does not like teaching logic, and prefers syntax, but she is not allowed to teach both. This year, it is likely that she needs to teach logic, and if so, she prefers to do so in the morning, as she prefers to have all her teaching in the morning.
a. Natasha hopes that she'll teach logic in the MORning. TRUE
b. Natasha hopes that she'll teach LOgic in the morning. FALSE

[4] We could include all non-trivial stronger propositions in the denotations, as in certain versions of Inquisitive Semantics, but such elaborate structure is unnecessary for the purposes of this paper.

[5] Following [13], we assume that exhaustive readings of embedded questions result from (optional) operators. See [30].

[6] Following [17], we write presuppositions after the colon in lambda terms.

Similar observations suggest preferential predicates are generally focus sensitive. [27] provides the following example for *surprise*:

(11) CONTEXT: Lisa knew that syntax was going to be taught. She expected syntax to be taught by John, since he is the best syntactician around. Also, she expected syntax to be taught on Mondays, since that is the rule.

a. It surprised Lisa that John taught syntax on TUESdays. TRUE
b. It surprised Lisa that JOHN taught syntax on Tuesdays. FALSE

These observation show that the alternatives that are compared in the semantics of preferential predicates are partly determined by the focus structure.

The degree-based semantics for preferentials by [27] builds on this insight, and treats the focus structure of the complement as providing the *comparison class* against which the subject's preferences are compared. Concretely, assuming the Roothian focus semantics ([28]), we take the context to provide a set of alternatives C, which preferential predicates refer to.[7] For example, the semantics for *be happy* looks like (12) with the auxiliary definitions for functions $\texttt{Pref}$ and θ in (13):[8]

(12) $[\![\textsf{be happy}_C]\!]^w$
$= \lambda p_{\langle s,t\rangle}.\lambda x\colon\ p(w) \wedge B_w(x,p) \wedge p \in C.\ \texttt{Pref}_w(x,p) > \theta(\{\texttt{Pref}_w(x,p') \mid p' \in C\})$

(13) a. $\texttt{Pref}_w(x,p) :=$ the degree to which x prefers p at w
b. $\theta(\{d_1, d_2, ..., d_n\}) := \sum_{i=1}^{n} d_i/n$

In prose, x *is happy that* p presupposes that p is true, x believes that p, and p is a member of the focus alternatives C,[9] and asserts that the degree to which x prefers p at w is greater than the *average* degree of $x's$ preferences for alternatives in C.

Note that (12) assumes that *be happy* semantically selects for a proposition. To reformulate the analysis to fit the uniform approach to clausal embedding introduced in the previous section, we make the predicate select for a set of propositions and relate the subject and the set using (13) via existential quantification:[10]

(14) $[\![\textsf{be happy}_C]\!]^w = \lambda Q_{\widehat{\langle s,t\rangle}}.$
$\lambda x\colon\ \exists p \in Q[p(w) \wedge B_w(x,p) \wedge p \in C].\ \exists p'' \in Q \left[\begin{array}{c} p''(w) \wedge B_w(x,p'') \wedge p'' \in C \wedge \\ \texttt{Pref}(x,p'') > \theta(\{\texttt{Pref}(x,p') \mid p' \in C\}) \end{array} \right]$

Let us see how (14) works with concrete interrogative and declarative complements. First, following [4], we take *wh*-items to be necessarily focused. Given this, in our semantics, the focus-semantic value of a *wh*-complement turns out to be equivalent to its ordinary-semantic value, as in (15).[11] Letting $\mathcal{Q}$ be the focus/ordinary semantic value of the interrogative complement,

[7] We assume for the sake of exposition that focus association with preferential predicates is conventional (in the sense of [3]), but nothing crucial hinges on this. See [27] for discussion. Also to avoid clutter, we conflate variables in the object language and meta-language.

[8] The formulation in (i) uses a *measure function* that returns degrees from individuals/propositions à la [20] instead of relations between degrees and individuals/propositions used in [27]. This is because of presentational reasons (the former formulation results in shorter formulae) and nothing hinges on this technical choice.

[9] As [27] argues, the last presupposition is an instance of a presupposition existing in degree constructions in general, that the comparison class includes the comparison term.

[10] In (i), to avoid the 'binding problem' concerning the existential quantifications in the presupposition and the assertion, the conditions in the presupposition are 'repeated' in the scope of the existential quantification in the assertion. See [29] for a similar solution to the binding problem in the domain of question-embedding.

[11] Whether this equivalence can be maintained for singular-*which* questions is unclear. The question denota-

be happy with an interrogative complement can be analyzed as in (16):

(15) $\mathcal{Q} := [\![\text{who jumped}]\!]^w = [\![[\text{who}]_F \text{ jumped}]\!]^f$

(16) $[\![\text{John is happy}_C\ [[\text{who}]_F \text{ jumped}]{\sim}C]\!]^w$ is

- defined only if $\exists p{\in}\mathcal{Q}[p(w) \wedge B_w(j,p) \wedge p \in C]$; if defined,
- true iff $\exists p''{\in}\mathcal{Q}\left[\begin{array}{c} p''(w) \wedge B_w(j,p'') \wedge p'' \in C \\ \texttt{Pref}(j,p'') > \theta(\{\texttt{Pref}(j,p') \mid p' \in C\}) \end{array}\right]$

Given the definition of the $\sim$-operator in (17) ([27]; cf. [28]), C in (16) is constrained as in (18).

(17) a. $[\![\alpha{\sim}C]\!]^o$ is defined only if $C \subseteq [\![\alpha]\!]^f$; if defined, $[\![\alpha{\sim}C]\!]^o = [\![\alpha]\!]^o$
b. $[\![\alpha{\sim}C]\!]^f$ is defined only if $C \subseteq [\![\alpha]\!]^f$; if defined, $[\![\alpha{\sim}C]\!]^f = [\![\alpha]\!]^f$

(18) $C \subseteq [\![\text{who jumped}]\!]^f = \mathcal{Q}$

All in all, (16) presupposes that there is a true answer of $\mathcal{Q}$ which John believes, and asserts that a true answer of $\mathcal{Q}$ which John believes is such that he prefers it to a greater extent than his average preferences for the alternatives in C, which in turn is a subset of $\mathcal{Q}$.

Next, a declarative-embedding sentence would be analyzed as in (19), with the variable C constrained by the focus structure as in (20). (Here, we let $A := \lambda w.J_w(a)$.)

(19) $[\![\text{John is happy}_C \text{ that } [[\text{Alice}]_F \text{ jumped}]{\sim}C]\!]^w$ is

- defined only if $\exists p{\in}\{A\}[p(w) \wedge B_w(j,p) \wedge p \in C]$
  $\equiv A(w) \wedge B_w(j,A) \wedge A \in C$; if defined
- true iff $\exists p''{\in}\{A\}\left[\begin{array}{c} p''(w) \wedge B_w(j,p'') \wedge p'' \in C \wedge \\ \texttt{Pref}(j,p'') > \theta(\{\texttt{Pref}(j,p') \mid p' \in C\}) \end{array}\right]$
  $\equiv A(w) \wedge B_w(j,A) \wedge A \in C \wedge \texttt{Pref}(j,A) > \theta(\{\texttt{Pref}(j,p') \mid p' \in C\})$

(20) $C \subseteq [\![\text{that [Alice]}_F \text{ jumped}]\!]^f = \mathcal{Q}$

That is, (19) presupposes that Alice jumped and that John believes that Alice danced, and asserts that John prefers Alice's jumping to a greater extent than his preferences for the alternatives in C, which again is constrained by $\mathcal{Q}$.

Thus, the degree-based analysis provides a straightforward account of both declarative and interrogative-complementation under veridical preferentials. [27] shows that the degree-based analysis enables an attractive account of two puzzles concerning veridical preferentials: (i) incompatibility with *whether*-complements and (ii) (typical) incompatibility with strongly-exhaustive embedded questions. Another virtue of the degree-based analysis is that (with a suitable syntax-semantics assumptions) it can account for the behavior of preferential predicates as gradable predicates, as in their occurrence in comparatives:

(21) a. Andrew is happier that Alice jumped than Bill is.
b. Ben liked/hated that Alice jumped more than Bill did.

3.3 Deriving the Anti-rogativity of Non-veridical Preferentials

Building on the semantics for veridical preferentials in the previous section, we propose the semantics of non-veridical preferential, such as *hope*, as follows:

tion only contains 'singular' answers [10] while the focus alternatives may contain 'plural' answers depending on the analysis of the focus value of *which*-NPs. We thank Henriette de Swart for pointing out this potential issue.

(22) $[\![\mathsf{hope}_C]\!]^w$
$= \lambda Q_{\widehat{\langle s,t\rangle}}.\lambda x : \exists p \in Q[p \in C].\ \exists p'' \in Q[p'' \in C \wedge \mathtt{Pref}(x, p'') > \theta(\{\mathtt{Pref}(x, p') \mid p' \in C\})]$

In contrast to the veridical preferential *be happy* in (14), which requires that the preferred answer is *true* and is *believed by the subject*, the non-veridical preferential *hope* in (22) lacks such requirements. The body of the function simply states that there is an answer (which is also a member of C) that the subject prefers to a greater extent than the average given C.

With a declarative complement, (22) derives the meaning that the subject prefers the proposition denoted by the complement to a greater degree than the average given focus alternatives:

(23) $[\![\mathsf{John\ hopes}_C\ \mathsf{that\ [[Alice]}_F\ \mathsf{jumped]}{\sim}C]\!]^w$ is
- defined only if $A \in C$; if defined,
- true iff $\mathtt{Pref}(j, A) > \theta(\{\mathtt{Pref}(j, p') \mid p' \in C\})$

On the other hand, the meaning predicted for (22) with an interrogative complement, exemplified in (24), turns out to be systematically trivial, assuming an additional presupposition triggered by θ, given in (25).

(24) $[\![\mathsf{John\ hopes}_C\ \mathsf{[[who]}_F\ \mathsf{jumped]}{\sim}C]\!]^w = 1$ iff
- defined only if $\exists p \in \mathcal{Q}[p \in C]$; if defined,
- true iff $\exists p'' \in \mathcal{Q}[p'' \in C \wedge \mathtt{Pref}(j, p'') > \theta(\{\mathtt{Pref}(j, p') \mid p' \in C\})]$

(25) $\theta(\{d_1, d_2, ..., d_n\})$ is defined only if $\neg\exists d \forall d' \in \{d_1, d_2, ..., d_n\}[d = d']$

The presupposition states that the degrees in the comparison class cannot be all equal. In other words, in order for comparison to make sense, there has to be variability in the relevant degrees.[12] This amounts to (26) in the case of (24), i.e., that John's preferences vary.

(26) $\neg\exists d \forall d' \in \{\mathtt{Pref}(j, p') \mid p' \in C\}[d = d']$

Given the variability presupposition, (25) turns out to be necessarily true whenever it is defined. This is so since whenever John's preferences for the alternatives in C vary, there is always a proposition in $C \subseteq \mathcal{Q}$ which John prefers more than his average preference for C.

We follow [2, 12, 7] in assuming that systematic logical triviality leads to ungrammaticality. In particular, we assume the following principles from [12], where (27-a) is modified from the original to encompass presuppositional denotations.[13]

[12]The oddness of the following kind of example may empirically motivate the variability presupposition in degree constructions in general:

(i) #No Japanese semanticist is tall for a Japanese semanticist.

If it were not for the variability presupposition, (i) would be a felicitous sentence conveying that all Japanese semanticists are of the same height. However, we have to be inconclusive as to whether (i) counts as strong evidence for the variability presupposition, as the oddness of (i) could be explained differently (e.g., maxim of manner), as an anonymous reviewer for the Amsterdam Colloquium pointed out.

[13]We also need the definition of LOGICAL SKELETONS, which in turn requires the criteria for logical vocabularies. Following [33], [12] defines logical vocabularies in terms of *permutation invariance*. The present analysis can be made compatible with this view, by assuming (i) that the external argument of a predicate is introduced by a designated head (say v) [21] and that the clause-type operators !/? [8] are in charge of making sure whether the proposition-set denoted by the complement is singleton or multiple. In other words, (i) would be the logical skeleton for (23) and (24) where v, *hope* and !/? are the logical vocabularies.

(i) $[\ X_i\ [_{v'}\ v\ [_{\mathrm{VP}}\ \mathrm{hope}\ [_{\mathrm{CP}}\ !/?\ X_j]\ ]\ ]\ ]$

(27) a. An LF constituent a of type t is L-ANALYTIC iff a's logical skeleton receives the denotation 1 (or 0) under every variable assignment *when the denotation is defined*.
b. A sentence is ungrammatical if its Logical Form contains a L-analytic constituent.

Hence, the L-analyticity ensures that (24) is ungrammatical. On the other hand, even with the variability presupposition, the declarative-embedding variant is *not* L-analytic because of the singleton restriction on existential quantification. That is, (23) is contingent on whether John prefers A more than the average. This explains the anti-rogativity of non-veridical preferentials. Finally, veridical preferentials do not induce L-analyticity regardless of the complement clause type, due to the veridical restriction on existential quantification. The assertion of *be happy*+interrogative in (16) above is non-trivial (even with the variability presupposition) since its truth is contingent on whether John prefers a *true* answer. Similarly to the non-veridical case, *be happy*+declarative is non-trivial because of the singleton restriction.

4 Conclusions

In this paper, we have put forward a generalization that all non-veridical preferential predicates are anti-rogative, and provided a semantic explanation for this generalization using the uniform semantics of clausal-embedding predicates [8, 32, 30] and the degree-based semantics for preferential predicates [27]. The paper thus advances the currently active research into the semantic roots of selectional restrictions [9, 32, 31, 23].

References

[1] Pranav Anand and Valentine Hacquard. Epistemics and attitudes. *Semantics & Pragmatics*, 6(8):1–59, 2013.

[2] Jon Barwise and Robin Cooper. Generalized quantifiers and natural language. *Linguistics and Philosophy*, 4(2):159–219, 1981.

[3] David I. Beaver and Brady Z. Clark. *Sense and Sensitivity: How Focus Determines Meaning*. Wiley-Blackwell, West Sussex, 2008.

[4] Sigrid Beck. Intervention effects follow from focus interpretation. *Natural Language Semantics*, 14(1):1–56, 2006.

[5] Sigrid Beck and Hotze Rullmann. A flexible approach to exhaustivity in questions. *Natural Language Semantics*, 7(3):249–298, 1999.

[6] Dwight Bolinger. Post-posed main phrases: An English rule for the Romance subjunctive. *Canadian Journal of Linguistics*, 14(1):3–30, 1968.

[7] Gennaro Chierchia. *Logic in Grammar: Polarity, Free Choice, and Intervention*. Oxford University Press, Oxford, 2013.

[8] Ivano Ciardelli, Jeroen Groenendijk, and Floris Roelofsen. Inquisitive semantics: a new notion of meaning. *Language and Linguistics Compass*, 7(9):459–476, 2013.

[9] Ivano Ciardelli and Floris Roelofsen. Inquisitive dynamic epistemic logic. *Synthese*, 192(6):1643–1687, 2015.

[10] Veneeta Dayal. *Locality in WH Quantification: Questions and Relative Clauses in Hindi*. Kluwer Academic Publishers, Dordrecht, 1996.

[11] Paul Égré. Question-embedding and factivity. *Grazer Philosophische Studien*, 77(1):85–125, 2008.

[12] Jon Gajewski. L-Analyticity and Natural Language. Manuscript, MIT, 2002.

[13] Benjamin George. *Question Embeddng and the Semantics of Answers*. PhD thesis, University of California Los Angeles, 2011.

[14] Jane Grimshaw. Complement selection and the lexicon. *Linguistic Inquiry*, 10(2):279–326, 1979.

[15] Irene Heim. Presupposition projection and the semantics of attitude verbs. *Journal of Semantics*, 9(3):183–221, 1992.

[16] Irene Heim. Interrogative semantics and Karttunen's semantics for *know*. In *Proceedings of IATL 1*, pages 128–144, 1994.

[17] Irene Heim and Angelika Kratzer. *Semantics in Generative Grammar*. Blackwell, Oxford, 1998.

[18] Laurence Horn. *A Natural History of Negation*. Chicago University Press, Chicago, 1989.

[19] Lauri Karttunen. Syntax and semantics of questions. *Linguistics and Philosophy*, 1:3–44, 1977.

[20] Christopher Kennedy. Vagueness and grammar: the semantics of relative and absolute gradable adjectives. *Linguistics and Philosophy*, 30(1):1–45, 2007.

[21] Angelika Kratzer. Severing the external argument from its verb. In J. Rooryck and A. Zaring, editors, *Phrase Structure and the Lexicon*, pages 109–137. Kluwer, Dordrecht, 1996.

[22] Utpal Lahiri. *Questions and Answers in Embedded Contexts*. Oxford University Press, Oxford, 2002.

[23] Clemens Mayr. Predicting polar question embedding. In Robert Truswell, editor, *Proceedings of Sinn und Bedeutung 21*, Edinburgh, 2017. University of Edinburgh.

[24] David Pesetsky. Zero syntax, vol. 2: Infinitives. Ms., Massachusetts Institute of Technology.

[25] David Pesetsky. *Paths and Categories*. PhD thesis, Massachusetts Institute of Technology, 1982.

[26] Kyle Rawlins. About 'about'. In *Proceedings of SALT 23*, pages 336–357, 2013.

[27] Maribel Romero. *Surprise*-predicates, strong exhaustivity and alternative questions. In *Proceedings of SALT 25*, pages 225–245, 2015.

[28] Matts Rooth. A theory of focus interpretation. *Natural Language Semantics*, 1(1):75–116, 1992.

[29] Benjamin Spector and Paul Egré. A uniform semantics for embedded interrogatives: *an* answer, not necessarily *the* answer. *Synthese*, 192(6):1729–1784, 2015.

[30] Nadine Theiler, Floris Roelofsen, and Maria Aloni. A uniform semantics for declarative and interrogative complements. Manuscript, ILLC, University of Amsterdam, 2016.

[31] Nadine Theiler, Floris Roelofsen, and Maria Aloni. What's wrong with *believing whether*. In *Proceedings of SALT 27*, pages 248–265, 2017.

[32] Wataru Uegaki. *Interpreting Questions under Attitudes*. PhD thesis, Massachusetts Institute of Technology, 2015.

[33] Johan van Benthem. Logical constants across types. *Notre Dame Journal of Formal Logic*, 30(3):315–342, 1989.

[34] Elisabeth Villalta. Mood and gradability: an investigation of the subjunctive mood in Spanish. *Linguistics and Philosophy*, 31(4):467–522, 2008.

[35] Aaron Steven White and Kyle Rawlins. A computational model of S-selection. In *Proceedings of SALT 26*, pages 641–663, 2016.

[36] Richard Zuber. Semantic restrictions on certain complementizers. In *Proceedings of the 13th International Congress of Linguists*, 1982.

QUDs, brevity, and the asymmetry of alternatives

Matthijs Westera

Universitat Pompeu Fabra
matthijs.westera@gmail.com

Abstract

Exhaustivity is typically explained in terms of the exclusion of unmentioned alternatives. For this to work, the set of alternatives must be *asymmetrical*, lest both a proposition and its negation get excluded, yielding a contradiction (the *Symmetry Problem*). Since exhaustivity is regularly observed, these alternative sets must tend to be asymmetrical, and this requires an explanation. Existing explanations are based on considerations of brevity, but these run into certain problems. A new solution is proposed, explaining the asymmetry of alternatives in terms of the fact that discourse strategies with asymmetrical questions under discussion (QUDs) are favored because they allow part of the answer to be communicated implicitly, namely as an exhaustivity implicature.

1 Introduction

In (1), B's answer with falling intonation can be interpreted exhaustively:

(1) A: Who (of your friends John, Mary, Bill, Sue and Chris) was at the party?
B: John and Bill were there. → not Mary, not Sue, not Chris

Exhaustivity is typically explained in terms of the exclusion of unmentioned (and non-entailed) alternatives. For this to work, the set of alternatives must be *asymmetrical*, lest both a proposition and its negation get excluded, which would yield a contradiction. For instance, in (1) the set must contain only people's presences, not people's absences, for excluding both Mary's presence and Mary's absence yields a contradiction. Since exhaustivity is regularly observed, these alternative sets must tend to be asymmetrical, and this requires an explanation. This was pointed out by Kroch [21] (and subsequently discussed in [12, 25], among others) and it is currently known as the *Symmetry Problem* (attributed to MIT course notes by Heim and von Fintel).

Gazdar [9] proposed a potential solution to the Symmetry Problem in terms of *scales*: lexical entries would be associated with certain intrinsically asymmetrical scales of alternatives [14]. However, Russell [29] points out that scales aren't really a solution to the Symmetry Problem unless one explains why scales are the way they are, and why they should be what drives exhaustivity; and Geurts [10] notes that there is only very little explicit reflection on what scales are supposed to be. One option is to conceive of scales as indirect representations of *what is typically relevant* given that a certain lexical expression is used (following [22], [10] and presumably [14]). Another is to conceive of scales as representations not of what is typically relevant but of what is *actually* relevant for a given utterance (following, I believe, [13] and [24]; in this role scales are also called "Hirschberg scales" or "ad hoc scales" [18]). But regardless, as several authors note, scales don't explain the asymmetry they describe (e.g., [15, 29, 10]).

Previous explanations for the asymmetry of alternative sets have relied on considerations of *brevity* (e.g., [26, 1, 16, 22]). To illustrate:

(2) A: Were (all of) your friends at the party?
B: *Some* of them were there. → not all

To explain the exhaustivity, the alternative set must contain "all (of them were there)", but not its mirror image "some but not all". The brevity-based approach proposes that this is because "some but not all" is too cumbersome to express, unlike "all" or "some". For (1) a similar explanation may be given by assuming that "weren't" is significantly more cumbersome to express than "were". This approach faces some challenges that will be discussed in more detail in section 3. One of these is the possibility of exhaustivity on negative answers (again depending on intonation):

(3) B: (Of your friends,) *Mary* and *Sue* weren't there. → the others were there.

For the brevity-based explanation to explain this, "weren't there" must now be *less* complex than "were there", the opposite of what was required above.[1]

This paper proposes that challenges for existing brevity-based accounts, like (3), stem from relying on the wrong kind of 'brevity'. A division of pragmatic labor exists between choosing conversational goals and selecting the means for pursuing them, and, crucially, both choices may be guided by considerations of brevity. Previous approaches have concentrated exclusively on the brevity of one utterance compared to alternative utterances that would address the same goals; this paper proposes to consider also the brevity benefits of pursuing one set of goals rather than another, or, as I will say following much work in pragmatics, the brevity benefits of pursuing one Question Under Discussion (QUD) rather than another (e.g., [27]). The explanation for the asymmetry of alternative sets is then, in a nutshell, that pursuing asymmetrical QUDs rather than their symmetrical counterparts favors brevity.

Although considerations of brevity and notions like conversational goals or QUDs are fundamentally pragmatic, and I will assume a pragmatic source of exhaustivity in what follows, the proposed explanation of the asymmetry of alternative sets is intended to apply independently of whether exhaustivity is derived pragmatically or as part of some linguistic convention. This is because the explanation can be understood either as a synchronic (speaker-level) rationalization for pursuing asymmetrical QUDs, or as a diachronic (population-level) explanation for asymmetrical lexical scales, when these are conceived of as conventionalizations of *typical* QUDs (the perspective on scales taken in, e.g., [10, 22]). Previous brevity-based explanations, and most pragmatic explanations, likewise permit both interpretations.[2]

2 The solution: splitting a symmetrical QUD

Suppose for the sake of argument that A's interests in (1) are symmetrical, and that A's question introduces a symmetrical QUD, e.g.:

$$\text{QUD} = \{Pj, Pm, Pb, Ps, Pc, \overline{Pj}, \overline{Pm}, \overline{Pb}, \overline{Ps}, \overline{Pc}\}$$

(Additional constraints like closure under union and intersection may also be assumed, but will not matter in what follows.) In fact I think that A's interrogative in (1) does not make a particularly strong case for the QUD being symmetrical. But I also think that one can add "and who wasn't there?" to A's interrogative, which is more suggestive of a symmetrical QUD, without this making an exhaustive interpretation on B's response impossible. So, for the sake of

[1] Katzir's [19] 'complexity' does achieve this, but as a consequence it cannot be understood as implementing a global, pragmatic preference for brevity – and Katzir notes it is not intended as such. See also footnote 4.

[2] The possibility that asymmetrical QUDs have conventionalized as scales doesn't tell us much about *how* exhaustivity would arise conventionally, e.g., by means of certain constraints on the use of invisible operators [7] or otherwise.

argument, let us assume a symmetrical QUD in (1). In addition, let us assume that B's response in example (1) targets the same, symmetrical QUD supposedly introduced by A's question.

In addition, I will assume that exhaustivity follows in some way from compliance with the maxims, whether through considerations of Quantity [30], or, if one pleases, through an operator or other type of linguistic convention plus (often left implicit by proponents) Manner and Quality. If indeed exhaustivity follows in some way from the conversational maxims, then the assumption that B's response in example (1) would comply with the maxims relative to the symmetrical QUD leads to a contradiction – this is the Symmetry Problem. Put differently: B's response cannot address such a QUD while complying with the maxims. Although speakers may in principle violate maxims, namely in case of a clash, as Grice [11] noted they must not do so silently, lest they be liable to mislead; and I have proposed elsewhere [31] that maxim violations are signaled prosodically, by a final rising contour (or in written text by, e.g., "..."). In the absence of such cues, as in example (1), only one conclusion is possible: contrary perhaps to appearances, B's answer must be aimed at a different QUD, i.e., different from the symmetrical QUD supposedly introduced by A.

This conclusion, that speaker B in (1) cannot be addressing the symmetrical QUD, is in a way just a restatement of the Symmetry Problem. But this restatement could also be the first part of the solution, provided we can answer the important issues this raises:

(i) Which QUD is (or which QUDs are) addressed by speaker B in (1), if not the symmetrical QUD supposedly introduced by speaker A?

(ii) Why was this a rational choice of QUD for B?

(iii) How can an addressee (e.g., speaker A) figure this out, accommodate the new QUD(s) and compute the right inferences?

To complete the explanation, then, I will try to answer each of these questions in a thorough way.

Question (i): Which QUDs are addressed by B? I propose that for some reason (see question (ii) below) speaker B in (1) decided to split the prior QUD, if it was indeed symmetrical, into two asymmetrical QUDs, which I will denote by QUD^+ and QUD^-:

$$\text{QUD}^+ = \{Pj, Pm, Pb, Ps, Pc\} \qquad \text{QUD}^- = \{\overline{Pj}, \overline{Pm}, \overline{Pb}, \overline{Ps}, \overline{Pc}\}$$

It should be uncontroversial that an utterance that addresses multiple QUDs should convey an appropriate communicative intent for each QUD, i.e., an intent which complies with the maxims relative to that QUD. This explains why B's response in (1) would be fine with the assumed QUDs, in the following way:

1. B's primary (asserted/explicit) intent is that John and Bill were at the party, which can safely comply with the maxims relative to QUD^+;
2. because QUD^+ is asymmetrical, compliance with the maxims of the primary intent relative to this QUD safely implies exhaustivity, i.e., that according to the speaker Mary, Sue and Chris were absent;
3. the exhaustivity implication in turn enables the clear communication of a secondary intent, i.e., an conversational implicature, namely that Mary, Sue and Chris were absent;[3]

[3]The distinction between implication and implicature is important [2]: what is implied is not necessarily implicated (meant), and what is implicated is not necessarily implied to be true (but, typically, implied to be held true by the speaker).

4. the secondary intent can safely comply with the maxims relative to the other asymmetircal QUD^{-}.

That is, instead of addressing "who was there and who wasn't there?", for some reason speaker B decided to address only the positive half explicitly, enabling B to address the negative half implicitly by means of an exhaustivity implicature.

Some authors may disagree with my invocation of asymmetrical QUDs, for they may conceive of the Symmetry Problem as a deeper problem, due to a speaker's interests or 'relevance' being necessarily symmetrical (e.g., [4]; [8]; [7]). There are no good arguments for this position in the literature, and it contrasts with the better-argued stance of Horn [15] and Leech [23] that speakers tend to be much more interested in what there is than in what there isn't – an instance of what Horn calls the *Asymmetry Thesis* [17]. Moreover, in life we rely on default assumptions all the time, only the negations of which will be worth asserting (because only the negations of our default assumptions have the potential to change our plans and behaviors). But all of this is arguably irrelevant, because even if a speaker's interests happen to be symmetrical now and then, that doesn't mean that the QUDs must in that case also be symmetrical. The reason is that the choice of QUDs depends not just on what information is deemed interesting/relevant, but also on what the best discourse strategy is for getting this information into the common ground in an clear, orderly and efficient way [27]. It may well be rational to organize one's occasionally symmetrical interests into asymmetrical QUDs – which brings us to the following question.

Question (ii): Why was this QUD-split rational? Splitting a QUD into two is an ordinary *discourse strategy* [27], and one which in this particular situation offers a substantial benefit over simply addressing the original, symmetrical QUD. The benefit is that addressing an asymmetrical QUD enables an exhaustivity implicature, unlike the original symmetrical QUD (that's the Symmetry Problem) and that the exhaustivity implicature enables the speaker to address half of the original QUD (namely, the other asymmetircal half) implicitly, which greatly benefits brevity. In a sense, the symmetry problem solves itself: an asymmetrical QUD is favored precisely because the symmetrical QUD prevents an exhaustivity implicature. Interestingly this explanation, unlike existing brevity-based accounts, generalizes to exhaustivity on negative answers like (3), repeated here:

(3) B: (Of your friends,) *Mary* and *Sue* weren't there. → the others were there.

There is no reason why a speaker shouldn't decide to address the negative QUD explicitly, explicating who was *absent* and implicating that the others were *present*.

Now, this explanation doesn't mean that this kind of QUD-split is always appropriate. For instance, if speaker A is ticking boxes on a checklist of individuals, it might be better for B to address the entire original QUD explicitly, and in the precise order of the checklist:

(4) B: John was there, Mary wasn't, Bill was, Sue wasn't, and Chris wasn't.

Moreover, addressing half of the original QUD implicitly may not be a good idea in cases where the domain of relevant individuals is not entirely clear, which would compromise the clear communication of the exhaustivity implicature. But in other circumstances B's decision in (1), to split the QUD, seems to be perfectly rational.

In general, the more people were present (and also the greater the domain of A's inquiry) the greater the brevity benefit of explicating only who was absent. But other factors will also play a role, such as which of the two properties, being present or being absent, is the most salient in the broader context, e.g., the negative answer in (3) seems particularly natural if B normally takes

attendance by writing down only the names of those who are absent. Other factors that may play a role are, for instance, which of the two predicates is the most readily lexically accessible, and which of the individuals' names B is more likely to mispronounce. But for present purposes such complications can be set aside, because the main brevity benefit is more general: relying on conversational implicature for conveying part of the answer benefits brevity regardless of the particular lexical entries involved, because an implicature is, indeed, implicit. In this regard my explanation is crucially different from existing brevity-based approaches to the Symmetry Problem, see section 3.

A final remark regarding this issue, before turning to question (iii). While the brevity benefit alone may explain why B chose to split the prior QUD into two halves, it does not explain why it should be split into a positive QUD and a negative QUD, i.e., QUD^+ and QUD^-, rather than, e.g.:

$$\text{QUD}_1 = \{Pj, \overline{Pm}, Pb, \overline{Ps}, Pc\} \qquad \text{QUD}_2 = \{\overline{Pj}, Pm, \overline{Pb}, Ps, \overline{Pc}\}$$

After all, this split could have offered, depending on who was actually at the party, a similar brevity benefit as the split into QUD^+ and QUD^- (though with different predicted exhaustivity implicatures). One reason why the above split may be dispreferred is that the resulting QUDs are more complex: the propositions in QUD^+ and QUD^- vary only along a single dimension, i.e., the individual, whereas the propositions in QUD_1 and QUD_2 vary along two dimensions, i.e., the individual, and whether they were absent or present – and they vary in a rather unpredictable way, including some but not all combinations of individual and absence/presence. This added complexity would compromise clarity: an addressee may not be able to figure out which of many possible asymmetrical-but-mixed QUDs the speaker may be addressing (also if we take prosodic focus into account, see below). I take this to explain why the QUD-split must be as assumed, i.e., into QUD^+ and QUD^-.

Question (iii): How could an addressee figure this out? The main answer to this question is that B's response in (1) *must* be aimed at a different QUD, because it would have violated a maxim relative to the symmetrical QUD – and addressees should recognize this. Which different QUD(s) speaker B may be addressing is constrained, in turn, by the notion of discourse strategy: it must be some combination of QUDs that together cover the original one, and we have already explained why the assumed split into a positive and a negative QUD is favored over the more arbitrary mixes.

Besides these general pragmatic considerations, in spoken language addressees may of course also rely on *prosodic focus* for identifying the QUD. For B's response in (1) to imply exhaustivity, it should have a pitch accent on the individuals' names but not on the predicates. Let us assume that accent placement reflects, through focus structure in the usual manner (e.g., [28, 3]), only the primary QUD, i.e., the QUD that is explicitly addressed. The focus structure of (1) will then help an addressee to figure out that the primary QUD of B's response is the asymmetrical, positive one. In contrast, if B had been addressing the symmetrical QUD, or a strange mixture like QUD_1 or QUD_2 above, B should have used either broad focus (i.e., on the entire sentence, which would normally entail an accent at least on the predicates) or multiple foci, i.e., both the individuals' names and the predicates. To illustrate, example (4) does address the symmetrical QUD, and indeed an intonation contour with accents on both the names and the predicates seems the most natural there.

Summing up, pragmatic accounts of exhaustivity that predict a contradiction relative to a symmetrical QUD, enable a solution to the Symmetry Problem precisely because this contradiction means that a QUD-shift must have taken place. The QUD-shift can be understood in

terms of a rational discourse strategy, namely, that of splitting a symmetrical QUD into two asymmetrical QUDs, which offers a brevity benefit by virtue of enabling part of the answer to be communicated via exhaustivity implicature.

3 A closer look at previous brevity-based approaches

Approaches based on brevity would attempt to solve the Symmetry Problem in cases like (1) by assuming that "John wasn't there" is a more complex expression than "John was there", and likewise for the other individuals. This would provide speakers with an excuse for omitting Mary's absence but not for omitting Mary's presence, thus breaking the problematic symmetry (e.g., [22]). Several authors have tried to define an appropriate notion of brevity/complexity, for instance in terms of number of syllables [26] or degree of lexicalization [1, 16].[4] But this type of approach runs into several problems.

First, although it seems true that "wasn't" is more complex than "was", what this approach crucially need to assume is that the purported difference in complexity would be *sufficiently large* for it to matter. That is, the difference should be sufficiently large to provide speakers with an excuse for not mentioning certain relevant propositions, *and* for this omission to not cause any confusion among the addressees. Moreover, this would have to hold even if in other regards the speaker appears not to care too much about brevity, e.g.:

(5) Well, that is a most interesting question indeed, and I am delighted to be able to assist. Of your dear friends, John was there, and Mary was there.

After all, this seems to imply exhaustivity in the same way. Furthermore, this approach would have to assume a similarly significant brevity difference even between "was absent" and "was present", by which "wasn't there" and "was there" can be replaced in the relevant examples without changing the exhaustivity implications.

But even (or especially) if it can be shown that "wasn't" (or "absent") is *sufficiently* more complex than "was" (or "present"), there is still the problem of exhaustivity on negative answers like (3). In order for previous brevity-based accounts to explain this, "was" would have to be *more* complex than "wasn't" – the converse of what is needed for (1). What example (3) shows is that a brevity-based solution to the symmetry problem that feeds only on intrinsic properties of particular lexical entries is inadequate; rather, there must be a contextual parameter of, say, "mentionworthiness", that has nothing necessarily to do with intrinsic brevity or complexity. A similar criticism is voiced by Matsumoto [25], based on cases where a simple expression and a more complex expression are used together, e.g.:

(6) B: It was warm today, and a little bit more than warm yesterday.

Matsumoto observes that the utterance implies that it was not a little bit more than warm today, despite this being expressible only by a more complex utterance. In response, Katzir [19] proposes that sometimes complex expressions can be used in spite of their complexity, and that one can find out whether complex expressions can be used by checking whether the

[4] Katzir [19] tries to filter something like relevance in terms of a measure of grammatical complexity, in a way that would superficially seem to belong in the same strand as the aforementioned approaches. However, Katzir does not intend this to be part of a pragmatic explanation, and indeed it is difficult to see how it could be. Katzir's measure of grammatical complexity is defined in terms of whether certain permissible substitutions enable one to transform one sentence into another. A consequence of this is that which of two expressions counts as more complex in Katzir's sense can depend on which expression was actually uttered. Although we can see the appeal of this proposal within the otherwise unappealing grammatical approach to exhaustivity, it does not follow from a global pragmatic preference for brevity.

utterance itself contains such a complex expression somewhere. (Lassiter [22] presents a similar view in defense of the brevity-based approach.) This is of course true, but it doesn't explain why a particular context would be such that the more complex expression could be used to begin with. What it shows is that brevity-based approaches must invoke a contextual parameter of "mentionworthiness" that is at least in part independent of considerations of intrinsic brevity or complexity. (Neither author, to my awareness, considers exhaustivity on negative answers like (3).)

Once the need for a contextual "mentionworthiness" parameter is acknowledged, which given example (3) and arguably (5) cannot have anything necessarily to do with intrinsic lexical brevity or complexity, we may as well call the set of propositions that are worth mentioning a "QUD" and get rid of whatever symmetrical notion of relevance was used before (we can always obtain it by closing the notion of mentionworthiness or QUD under negation, should we find a need for it). The resulting picture is essentially what section 2 supports, by explaining why a speaker would choose to address an asymmetrical QUD despite symmetrical interests.

Recall that intrinsic lexical brevity or complexity did not play a role in the explanation I have proposed. Rather, it relied on the obvious brevity benefit of conversationally implicating part of the answer, which obtains in (1) and (3) alike. Another important difference is that, in my explanation, considerations of brevity are not strictly necessary for an audience to be able to identify the exhaustivity implicature: brevity may help explain *why* a speaker chose to address an asymmetrical QUD, but *that* the speaker did so will be evident regardless, from the fact that the utterance would have violated a maxim otherwise (as well as from prosodic focus). In contrast, according to previous brevity-based approaches, the audience would not be able to understand the exhaustivity implicature except through taking brevity into account. That such relatively tiny brevity differences would play such a central role does not seem plausible [5].

4 Conclusion and discussion

This paper argued that even if a speaker's interests are symmetrical – whether in general, which is questionable, or occasionally – it will often be rational for the speaker to organize the propositions of interest into asymmetrical QUDs – and the latter are what matters for exhaustivity. That is, it will be rational to split a symmetrical QUD into two asymmetrical halves because only an asymmetrical QUD permits conveying part of the answer implicitly, i.e., as an exhaustivity implicature, which favors brevity. This is a new explanation for why the alternative sets on which exhaustivity relies tend to be asymmetrical. With previous brevity-based explanations this proposal shares that considerations of brevity have some role to play, though with important differences in the kind of role: the proposed explanation relies only on the general fact that conversational implicature benefits brevity, regardless of the particular lexical entries involved.

I did not discuss the grammatical approach to exhaustivity. Chierchia et al. [7] list the Symmetry Problem as an argument against pragmatic theories of exhaustivity, and in favor of the grammatical approach. This paper shows that this argument falls short: a pragmatic explanation for asymmetrical alternative sets is available. Within the grammatical approach itself, symmetry is sometimes relied upon in order to block certain undesirable exhaustivity inferences. For instance, in order to block the "not all"-inference of the disjunction "some or all", local exhaustification of the first disjunct would render the two disjuncts mutually exclusive, which because of their symmetry would block subsequent global exhaustification (e.g., [6, 20]). Depending on how the sets of alternatives in the grammatical approach relate to

something like relevance or QUDs – and to my awareness there is no concensus in this regard – the arguments in the current paper may have some bearing on the grammatical approach as well. I leave an exploration of this relation to future work.

Zooming out a little, this paper highlighted an important *division of pragmatic labor*, namely between choosing certain QUDs to address and selecting appropriate communicative intents and expressions for doing so. More generally, this is a division between choosing/organizing one's goals and selecting the appropriate means to achieve them. Existing brevity-based approaches to the Symmetry Problem have concentrated on the means, by keeping the symmetrical QUD in place and comparing the brevity benefits only of different ways of addressing that QUD. My proposal, in contrast, considered the brevity benefits of a maneuver at the level of QUDs, i.e., of a certain discourse strategy, and this is what made it both explanatory and more successful at dealing with certain problems. I think it is essential for the field to keep this division of pragmatic labor in mind, and to explicate and motivate assumptions at both levels [32].

5 Acknowledgments

Many thanks to Floris Roelofsen and Jeroen Groenendijk for their comments on many iterations of this work. This project has received funding from the European Research Council (ERC) under the European Unions Horizon 2020 research and innovation programme (grant agreement No 715154). This paper reflects the authors' view only, and the EU is not responsible for any use that may be made of the information it contains.

References

[1] Jay David Atlas and Stephen C Levinson. It-clefts, informativeness and logical form: radical pragmatics (revised standard version). In P. Cole, editor, *Radical pragmatics*, pages 1–62. Academic Press, New York, 1981.

[2] Kent Bach. The top 10 misconceptions about implicature. In *Drawing the boundaries of meaning: Neo-Gricean studies in pragmatics and semantics in honor of Laurence R. Horn*, pages 21–30. John Benjamins Publishing Company, 2006.

[3] David Beaver and Brady Clark. *Sense and Sensitivity: How Focus Determines Meaning*. Number 12 in Explorations in Semantics. John Wiley & Sons, 2009.

[4] Eliza Block. Is the symmetry problem really a problem? Unpublished manuscript; retrieved from http://web.eecs.umich.edu/~rthomaso/lpw08/block.pdf, 2008.

[5] Robyn Carston. Relevance Theory, Grice and the neo-Griceans: a response to Laurence Horn's 'current issues in neo-Gricean pragmatics'. *Intercultural Pragmatics*, 2:303–319, 2005.

[6] Gennaro Chierchia, Danny Fox, and Benjamin Spector. Hurfords constraint and the theory of scalar implicatures. *Presuppositions and implicatures*, 60:47–62, 2009.

[7] Gennaro Chierchia, Danny Fox, and Benjamin Spector. The grammatical view of scalar implicatures and the relationship between semantics and pragmatics. In Claudia Maienborn, Paul Portner, and Klaus von Heusinger, editors, *Semantics: An International Handbook of Natural Language Meaning*, volume 2, pages 2297–2332. Mouton de Gruyter, 2012.

[8] Danny Fox and Roni Katzir. On the characterization of alternatives. *Natural Language Semantics*, 19(1):87–107, 2011.

[9] G. Gazdar. *Pragmatics: Implicature, Presupposition, and Logical Form*. Academic Press, New York, 1979.

[10] Bart Geurts. *Quantity Implicatures*. Cambridge University Press, 2011.

[11] H.P. Grice. *Studies in the Way of Words*. Harvard University Press, 1989.

[12] Jeroen Groenendijk and Martin Stokhof. *Studies on the Semantics of Questions and the Pragmatics of Answers*. PhD thesis, University of Amsterdam, 1984.

[13] Julia Hirschberg. *A Theory of Scalar Implicature*. PhD thesis, University of Pennsylvania, 1985.

[14] Laurence R. Horn. *On the Semantic Properties of Logical Operators in English*. PhD thesis, University of California Los Angeles, 1972.

[15] Laurence R. Horn. Lexical incorporation, implicature, and the least effort hypothesis. In Donka Farkas, Wesley M. Jacobsen, and Karol W. Todrys, editors, *CLS: Papers from the Parasession on the Lexicon*, pages 196–209. Chicago Linguistic Society, 1978.

[16] Laurence R. Horn. Towards a new taxonomy of pragmatic inference: Q-based and R-based implicatures. In D. Schiffrin, editor, *Meaning, Form, and Use in Context*, pages 11–42. Georgetown University Press, 1984.

[17] Laurence R. Horn. *A Natural History of Negation*. David Hume Series on Philosophy and Cognitive Science Reissues. CSLI Publications, Standford, 2001. First published in 1989.

[18] Yan Huang. *Pragmatics*. Oxford Textbooks in Linguistics. Oxford University Press, 2014.

[19] R. Katzir. Structurally-defined alternatives. *Linguistics and Philosophy*, 30(6):669–690, 2007.

[20] Roni Katzir and Raj Singh. Hurford disjunctions: embedded exhaustification and structural economy. In U. Etxeberria, A. Fălăuş, A. Irurtzun, and B. Leferman, editors, *Proceedings of Sinn und Bedeutung 18*. 2013.

[21] Anthony Kroch. Lexical and inferred meanings for some time adverbs. *Quarterly Progress Reports of the Research Laboratory of Electronics*, 104:260–267, 1972.

[22] Daniel Lassiter. Why symmetry is not a problem for a gricean theory of scalar implicature, 2010. Presented at Utterance Interpretation and Cognitive Models 3; retrieved from http://web.stanford.edu/~danlass/Lassiter-SI-UICM.pdf.

[23] Geoffrey Leech. Pragmatics and conversational rhetoric. In H. Parret, M. Sbisa, and J. Verschueren, editors, *Possibilities and limitations of pragmatics*, pages 413–442. John Benjamins, Amsterdam, 1981.

[24] Stephen C Levinson. *Pragmatics*. Cambridge textbooks in linguistics. Cambridge University Press, 1983.

[25] Yo Matsumoto. The conversational conditions on horn scales. *Linguistics and Philosophy*, 18:21–60, 1995.

[26] J.D. McCawley. Conversational implicature and the lexicon. In Peter Cole, editor, *Pragmatics*, number 9 in Syntax and Semantics, pages 245–259. Academic Press, 1978.

[27] Craige Roberts. Information structure in discourse. In J.H. Yoon and A. Kathol, editors, *OSU Working Papers in Linguistics*, volume 49, pages 91–136. Ohio State University, 1996.

[28] Mats Rooth. *Association with Focus*. PhD thesis, University of Massachusetts, Amherst, 1985.

[29] Benjamin Russell. Against grammatical computation of scalar implicatures. *Journal of Semantics*, 23:361–382, 2006.

[30] Westera. An attention-based explanation for some exhaustivity operators. In *Proceedings of Sinn und Bedeutung*. University of Edinburgh, 2017.

[31] Matthijs Westera. 'Attention, I'm violating a maxim!' A unifying account of the final rise. In Raquel Fernández and Amy Isard, editors, *Proceedings of the Seventeenth Workshop on the Semantics and Pragmatics of Dialogue (SemDial 17)*, 2013.

[32] Matthijs Westera. *Exhaustivity and intonation: a unified theory*. PhD thesis, submitted to ILLC, University of Amsterdam, 2017.

Widening Free Choice

Malte Willer

University of Chicago
willer@uchicago.edu

Abstract

Disjunctions scoping under possibility modals give rise to the free choice effect. The effect also arises if the disjunction takes wide scope over possibility modals; it is independent of the modal flavor at play (deontic, epistemic, and so on); and it arises even if disjunctions scope under or over necessity modals. At the same time, free choice effects disappear in the scope of negation or if the speaker signals ignorance or unwillingness to cooperate. I show how we can account for this wide variety of free choice observations without unwelcome side-effects in an update-based framework whose key innovations consist in (i) a refined test semantics for necessity modals and (ii) a generalized conception of narrow and wide scope free choice effects as arising from lexically or pragmatically generated prohibitions against the absurd state (an inconsistent information carrier) serving as an update relatum. The fact that some of these prohibitions are defeasible together with a binary semantics that distinguishes between positive and negative update relata accounts for free choice cancellation effects.

1 The Scope of Free Choice

It is a well-worn story that disjunctions scoping under possibility modals give rise to the *free choice* effect (Kamp 1973, von Wright 1968):

(1) You may take an apple or a pear.
 ⇝ a. You may take an apple.
 ⇝ b. You may take a pear.

(2) Mary might be in Rome or in Paris.
 ⇝ a. Mary might be in Rome.
 ⇝ b. Mary might be in Paris.

In both (1) and (2), the possibility of a disjunction seems to entail the possibility of each disjunct. This is unexpected on the standard analysis of modals and disjunction since the possibility of a disjunction is classically consistent with the impossibility of one of its disjuncts. The question addressed in this paper is how to do better.

One part of the explanandum is that free choice effects are not tied to some specific modal flavor: in (1) the modal is deontic, whereas in (2) it is epistemic. Another is that free choice effects may also arise if the disjunction takes *wide* as opposed to *narrow* scope with respect to the modal (see Kamp 1978 and also Geurts 2005; Simons 2005; Zimmermann 2000):

(3) You may take an apple, or you may take a pear.
 ⇝ a. You may take an apple.
 ⇝ b. You may take a pear.

(4) Mary might be in Rome, or she might be in Paris.
 ⇝ a. Mary might be in Rome.
 ⇝ b. Mary might be in Paris.

The examples in (3) and (4) seem to entail a conjunction of possibilities just as much as (1) and (2) do. A comprehensive story about the free choice effect must thus address both its narrow scope and its wide scope incarnations. It is, of course, tempting to reduce the latter to the former by appealing to Simons's (2005) proposal that across-the-board LF movement may transform

(3) into (1) and (4) into (2), respectively. But Starr (2016) observes that this approach faces an over-generation problem: if the relevant transformation could move a disjunction, it should also be able to move a conjunction—LF movement is a type-driven process, after all—and yet by everyone's agreement (5a) does not entail (5b).

(5) a. You may have an apple, and you may have a pear.
b. You may have an apple and a pear.

An important goal of this paper is to explain how a successful story about narrow free choice effects can also address their wide scope cousins without creating overgeneration problems.

Necessity modals form another part of the picture. It has often been observed that free choice effects arise if *must*s take scope over disjunction (see e.g. Aloni 2007). And again it looks as if the effect also arises if the disjunction takes wide scope, as the following examples indicate (assume that the *must* at play has a deontic flavor):

(6) You must wear a tuxedo or a black suit.
⇝ a. You may wear a tuxedo.
⇝ b. You may wear a black suit.

(7) You must wear a tuxedo, or you must wear a black suit.
⇝ a. You may wear a tuxedo.
⇝ b. You may wear a black suit.

None of this shows that possibility and necessity modals trigger free choice effects in exactly the same way: Aloni (2007), for instance, offers a pragmatic explanation of the entailments in (6) and a semantic explanation of the corresponding effect with possibility modals. The point is that any account that aims at explaining why disjunctions scoping under necessity modals trigger free choice effects should also have something to say about the inferences in (7).

The final piece of the puzzle is that free choice effects are in principle cancellable. Specifically, disjunction behaves classically if a disjunctive possibility is embedded under negation (Alonso-Ovalle 2006; Fox 2007):

(8) You may not take an apple or a pear.
⇝ a. You may not take an apple.
⇝ b. You may not take a pear.

(9) Mary cannot be in Rome or in Paris.
⇝ a. Mary cannot be in Rome.
⇝ b. Mary cannot be in Paris.

Free choice effects also disappear if the speaker signals ignorance (Kamp 1978) or uncooperativeness (Simons 2005). Neither (10a) nor (10b), for instance, suggest that one may choose between taking a pear and taking an apple:

(10) a. You may take an apple or a pear, but I don't know which.
b. You may take an apple or a pear, but I won't tell you which.

A comprehensive account of free choice effects not only has to explain why these effects occur, but also why they sometimes fail to occur.

Discussions of the free choice effect abound and a comprehensive review of what is currently on the market is better left to another day.[1] My goal here is to improve on the existing literature by telling a story that has something to say about all of the data discussed here so far.

[1] Pragmatic treatments of the free choice effect include Alonso-Ovalle 2006; Fox 2007; Franke 2011; Klinedinst 2007; Kratzer and Shimoyama 2002; Schulz 2005. Semantic approaches include Aher 2012; Aloni 2007, ms.; Barker 2010; Fusco 2015; Geurts 2005; Hawke and Steinert-Threlkeld ms., Roelofsen ms., Simons 2005; Starr 2016; Willer 2015, forthcoming; Zimmermann 2000.

2 Framework

The story told here draws inspiration from the relational analysis of modality and disjunction in Willer 2015, forthcoming, which again owes inspiration to seminal work in the dynamic and the inquisitive semantic tradition (in particular Aher 2012; Groenendijk and Roelofsen 2015; Veltman 1985, 1996). The target language $\mathcal{L}$ contains a set of sentential atoms $\mathcal{A} = \{p, q, ...\}$ and is closed under negation ($\neg$), conjunction ($\wedge$), disjunction ($\vee$), and the modal possibility and necessity operator ($\Diamond$,$\Box$), embellished with e, d, etc., as subscripts to distinguish between epistemic, deontic, and other modal flavors. The remaining connectives are defined as usual.

The proposal in Willer 2015, forthcoming treats semantic values as *update relations* and distinguishes between *positive* and *negative* updates. It makes sense of the initial observation that disjunctions scoping under possibility modals give rise to the free choice effect, and it also explains why the effect disappears if the disjunctive possibility occurs in the scope of a negation operator or in an ignorance/uncooperativeness context. But it remains silent on free choice effects that arise if the modal at play expresses a necessity or is outscoped by a disjunction. My goal here is to fill this lacuna.

The key idea idea of this paper is that—from an update-centric perspective anyway—free choice effects flow from a prohibition against updates that are inconsistent with some distinguished information carrier, and that such a prohibition may have a lexical source (e.g. the semantics for modals) but may also arise at the discourse level from general (and defeasible) assumptions of competence and cooperativeness. Together with a refined semantics for necessity modals—they test for compatibility in addition to testing for entailment—this idea provides the foundation for explaining the wide variety of free choice observations that were described earlier.

To get the revised test semantics for necessity modals off the ground, we have to explain how a state may support a disjunction of necessities such as (7) while rejecting both disjuncts if taken in isolation.[2] The key idea I wish to pursue here is that what makes the disjunction acceptable is a combination of two criteria: (i) each disjunct is supported given additional constraints on the modal domain and (ii) the constraints thus brought into play jointly exhaust logical space and hence amount to a trivial assumption. To make this idea precise, I let input contexts be pairs consisting of an inquisitive state—a set of consistent propositions, which I label here *alternatives*—and a proposition that may play the role of a restrictor on modal domains.

Definition: Input Contexts, States, Alternatives. w is a *possible world* iff $w\colon \mathcal{A} \mapsto \{0, 1\}$. W is the set of such w's, $\mathcal{P}(W)$ is the powerset of W. The function $[\![\cdot]\!]$ assigns to nonmodal formulas of $\mathcal{L}$ a *proposition* in the familiar fashion. An *input context* s_x is a pair consisting of an *inquisitive state* $s \subseteq \mathcal{P}(W)\backslash\{\varnothing\}$ and a *restrictor* $x \subseteq W$. Each element of an inquisitive state is an *alternative*. The *information* carried by a state s is the set of possible worlds compatible with it so that $\texttt{info}(s) = \{\bigcup \sigma\colon \sigma \in s\}$. S is the set of all states and I is the set of all input contexts. $\underline{\perp}$ represents any input context s_x with $s = \varnothing$ (any absurd context) while $\underline{\not\perp}$ represents any context s_x with $s \neq \varnothing$ (any non-absurd context).

Inquisitive states have informational content in the sense that they rule out certain ways the world could be. In addition, they encode this information as a set of alternatives (which do not have to be mutually exclusive). A context couples a state with a proposition that can play the

[2] Not everyone immediately agrees that (7) lends itself to a free choice interpretation, and indeed it is natural to interpret the speaker here as being uncertain about some particular dress code. Nonetheless it strikes me as uncontroversial that a free choice reading is available and that such a reading can be explicitly enforced, as in, e.g., "You must wear a tuxedo, or you must wear a black suit, it's up to you."

role of a restrictor on the modal domain—more on this momentarily—and the resulting pairs then play the role of update relata in our semantics.

Following a recent trend in the literature, I adopt a binary system that distinguishes between *positive* and *negative* update relations. Atomic sentences update as follows:

$$(\mathcal{A}) \quad \begin{array}{l} s_x[p]^+t_y \text{ iff } x = y \text{ and } t = \{\sigma \in s\colon \sigma \cap [\![p]\!] = \sigma\} \\ s_x[p]^-t_y \text{ iff } x = y \text{ and } t = \{\sigma \in s\colon \sigma \cap [\![p]\!] = \varnothing\} \end{array}$$

A positive update with a sentential atom p effectively eliminates from a state all alternatives that fail to entail p. A negative update with a sentential atom p eliminates from a state all alternatives that are compatible with p. This is just to say that a positive update with p removes all $\neg p$-worlds from the input context's informational content, while a negative update with p removes all p-worlds. Note that the restrictor is left alone in any case.

Not surprisingly, negative updates matter for the semantics of negation: we now say that an update with $\ulcorner\neg\phi\urcorner$ is a negative update with ϕ. The requirement that a negative update with $\ulcorner\neg\phi\urcorner$ is a positive update with ϕ delivers the law of double negation:

$$(\neg) \quad \begin{array}{l} s_x[\neg\phi]^+t_y \text{ iff } s_x[\phi]^-t_y \\ s_x[\neg\phi]^-t_y \text{ iff } s_x[\phi]^+t_y \end{array}$$

The current setup is, so far, only a complicated version of Update Semantics. The additional complexity, however, allows us to plug in an inquisitive analysis of disjunction as well as a sophisticated test analysis of modals. Here is the proposal for the former:

$$(\vee) \quad \begin{array}{l} s_x[\phi \vee \psi]^+t_y \text{ iff } s_x[\phi]^+t_y \text{ or } s_x[\psi]^+t_y \\ s_x[\phi \vee \psi]^-t_y \text{ iff } \exists u_z\colon s_x[\phi]^-u_z \text{ and } u_z[\psi]^-t_y \end{array}$$

A disjunction relates an input context to two potentially distinct output contexts: the result of updating with the first and the result of updating with the second disjunct.

Given some input context, a positive update with a conjunction $\ulcorner\phi \wedge \psi\urcorner$ proceeds via a positive update with ϕ and then via a positive update with ψ:

$$(\wedge) \quad \begin{array}{l} s_x[\phi \wedge \psi]^+t_y \text{ iff } \exists u_z\colon s_x[\phi]^+u_z \text{ and } u_z[\psi]^+t_y \\ s_x[\phi \wedge \psi]^-t_y \text{ iff } s_x[\phi]^-t_y \text{ or } s_x[\psi]^-t_y \end{array}$$

The rules for negative updates with disjunctions and conjunctions enforce the validity of De Morgan's Laws. This, I should add, is negotiable and it would be easy to change the negative entries if De Morgan's Laws turn out to be invalid (as argued by, for instance, Champollion et al. 2016, forthcoming). Let me instead move on to the update rules for modal expressions.

Modals are interpreted in light of a set of contextually provided modal selection functions, which map each state to a modal domain (another state) that can then be further restricted by the proposition provided by the input or output context:

Definition: Modal Selection Functions, Modal Domain Restrictions. A contextually provided *modal selection function* $f\colon S \mapsto S$ maps each state to a *modal domain* (another state). Given some state $s \in S$ and proposition p, we define the *restriction* of $f(s)$ with p as $f(s)\restriction_p = \{y \cap p\colon y \in f(s)\}\backslash\{\varnothing\}$.

Different modal flavors call for different modal selection functions and we will label them in the obvious way: e for epistemic, d for deontic, and so on. The restriction of a modal domain with

a proposition p proceeds as expected: we eliminate from each alternative in the state those worlds at which the proposition p is false and collect the results, leaving out the empty set.

We are now ready to present the proposal for modal expressions. Let us focus on the positive update rules first:

$$(\Diamond_f^+/\Box_f^+) \quad \begin{array}{l} s_x[\Diamond_f\phi]^+t_y \text{ iff } t = \{\sigma \in s\colon \langle f(s)_W, \underline{\bot}\rangle \notin [\phi]^+\} \text{ and } x = y \\ s_x[\Box_f\phi]^+t_y \text{ iff } t = \{\sigma \in s\colon \langle f(s)_W, \underline{\bot}\rangle \notin [\phi]^+\} \text{ and } \langle (f(s)\restriction_y)_W, \underline{\not\bot}\rangle \notin [\phi]^- \end{array}$$

Possibility modals are effectively tests à la Veltman 1996: for an input context to pass the test imposed by a positive update with '$\Diamond_f\phi$', the relevant modal domain must not be related to the absurd state via a positive update with the prejacent ϕ. Necessity modals test for consistency, too, but in addition require that the output context come with a restrictor enforcing the necessity of the prejacent in the modal domain: thus restricted, the modal domain is only related to the absurd state via a negative update with the prejacent.

It is straightforward to define the negative update rules for modals:

$$(\Diamond_f^-/\Box_f^-) \quad \begin{array}{l} s_x[\Diamond_f\phi]^-t_y \text{ iff } s_x[\Box_f\neg\phi]^+t_y \\ s_x[\Box_f\phi]^-t_y \text{ iff } s_x[\Diamond_f\neg\phi]^+t_y \end{array}$$

The negative entries effectively require that the possibility and the necessity modal be duals.

It remains to explain what it takes to update a state. As a preparation, define what it takes for a state and a restriction to be positively related to some state s via ϕ.

Definition: Positive Update Relata. Define a function $\Delta\colon S \times \mathcal{L} \mapsto S$ and a function $\Gamma\colon S \times \mathcal{L} \mapsto \mathcal{P}(W)$ as follows:

1. $\Delta(s,\phi) = \{t\colon \exists y.\, s_W[\phi]^+t_y \text{ and } y \neq \varnothing\}$
2. $\Gamma(s,\phi) = \{y\colon \exists t.\, s_W[\phi]^+t_y \text{ and } t \neq \varnothing\}$

An update of a state s with ϕ is the union of the states positively related to s via ϕ, provided that the union of restrictions positively related to s via ϕ amounts to a tautology—otherwise, the update returns the empty set. More precisely:

Definition: Updates on States, Support. An update function $\uparrow\colon S \mapsto S$ is defined as follows:

$$s \uparrow \phi \quad = \quad \begin{cases} \bigcup\Delta(s,\phi) & \text{if } \bigcup\Gamma(s,\phi) = W \\ \varnothing & \text{otherwise} \end{cases}$$

We say that s *supports* ϕ, $s \Vdash \phi$, iff $\texttt{info}(s \uparrow \phi) = \texttt{info}(s)$.

As we will see momentarily, this update procedure allows necessity modals to test for entailment, but in a roundabout way. A necessity modal effectively identifies which assumptions are required so that the modal domain entails the prejacent. The entailment test flows from the *general* requirement on updating that the restrictions thus brought into play amount to a trivial restriction of the modal domain.

Finally, we define entailment in the familiar dynamic fashion:

$$\phi_1, \ldots, \phi_n \textit{ entails } \psi,\ \phi_1, \ldots, \phi_n \models \psi, \text{ iff for all } s \in S,\ s \uparrow \phi_1 \ldots \uparrow \phi_n \Vdash \psi$$

A state supports ϕ just in case a positive update of s with ϕ does not add to the s's informational content. Entailment is guaranteed preservation of support. This setup will take care of all the narrow scope free choice data and—given suitable pragmatic supplementation—make sense of wide scope free choice as well.

3 Output

Narrow scope free choice effects arise both for possibility and for necessity modals:

Fact 1. $\Diamond_f(p \vee q) \vDash \Diamond_f p \wedge \Diamond_f q$ **Fact 2.** $\Box_f(p \vee q) \vDash \Diamond_f p \wedge \Diamond_f q$

The underlying observation here is that an update of s with "$\Diamond_f(p \vee q)$" or with "$\Box_f(p \vee q)$" results in the absurd state unless $\langle f(s)_W, \underline{\perp}\rangle \notin [p \vee q]^+$. But suppose that $f(s)$ fails to contain a p-entailing and a q-entailing alternative: then $[p]^+$ or $[q]^+$ *does* relate $f(s)_W$ to $\underline{\perp}$ and thus $\langle f(s)_W, \underline{\perp}\rangle \in [p \vee q]^+$ after all. So whenever $s \uparrow \Diamond_f(p \vee q) \neq \varnothing$ or $s \uparrow \Box_f(p \vee q) \neq \varnothing$, then $s \uparrow \Diamond_f p = s$ and $s \uparrow \Diamond_f q = s$. Note that the free choice inference arises regardless of modal flavor. Note furthermore that $\Diamond_f(p \vee q) \nvDash \Diamond_f(p \wedge q)$ since passing the test conditions under consideration does not require the presence of a $p \wedge q$-entailing alternative in $f(s)$.

We also account for the observation that embedding a disjunctive possibility or necessity under negation reverts disjunction to its classical behavior:

Fact 3. $\neg\Diamond_f(p \vee q) \vDash \neg\Diamond_f p \wedge \neg\Diamond_f q$ **Fact 4.** $\neg\Box_f(p \vee q) \vDash \neg\Box_f p \wedge \neg\Box_f q$

To see this, observe that our negative update rule for the possibility modal require that $s \uparrow \neg\Diamond_f(p \vee q) = s \uparrow \Box_f\neg(p \vee q)$. And if $f(s)$ were to include a p- or a q-entailing alternative, then $s \uparrow \Box_f\neg(p \vee q) = \varnothing$, which establishes Fact 3. Furthermore, our negative update rule for the necessity modal requires that $s \uparrow \neg\Box_f(p \vee q) = s \uparrow \Diamond_f\neg(p \vee q)$. The fact that $\Diamond_f\neg(p \vee q) \vDash \Diamond_f\neg p \wedge \Diamond_f\neg q$ establishes Fact 4 since $\ulcorner\Diamond_f\neg\phi\urcorner$ entails $\ulcorner\neg\Box_f\phi\urcorner$ by design.

It remains to comment on what the framework has to say about wide scope free choice effects. Start with the following observation: what drives narrow free choice effects is that modals require that their prejacents not relate the relevant modal domain to the absurd state. But of course such a prohibition need not only flow from the lexical semantics for modal expressions but also arises at the discourse level from general (and defeasible) assumptions of competence and cooperativeness. In fact, Stalnaker (1978) takes it a basic communicative principle that speakers should not assert what they presuppose to be false. We can capture this intuition in terms of the following QUALITY constraint, where s_c is the state capturing what is common ground in some context c:

Quality An assertion of ϕ in context c satisfies the quality constraint just in case $\varnothing \notin \Delta(s_c, \phi)$.

We may then define a pragmatically supplemented notion of entailment that evaluates arguments under the assumption that the QUALITY constraint is satisfied:

> $\phi_1, \ldots, \phi_n$ *pragmatically entails* ψ, $\phi_1, \ldots, \phi_n \gg \psi$, iff for all $s \in S$, if $\varnothing \notin \Delta(s, \phi_1)$ and $\ldots$ and $\varnothing \notin \Delta(s \uparrow \phi_1 \ldots \uparrow \phi_{n-1}, \phi_n)$, then $s \uparrow \phi_1 \ldots \uparrow \phi_n \Vdash \psi$

Every semantic entailment is also a pragmatic entailment, but the reverse need not hold.

Wide scope free choice effects are pragmatic entailments:

Fact 5. $\Diamond_f p \vee \Diamond_f q \gg \Diamond_f p \wedge \Diamond_f q$ **Fact 6.** $\Box_f p \vee \Box_f q \gg \Diamond_f p \wedge \Diamond_f q$

Assume that $f(s)$ fails to include a p- and a q-entailing alternative. Then either $[\Diamond_f p]^+$ or $[\Diamond_f q]^+$ relates s_W to $\underline{\perp}$ and thus $\varnothing \in \Delta(s, \Diamond_f p \vee \Diamond_f q)$, violating QUALITY. And for parallel reasons, either $[\Box_f p]^+$ or $[\Box_f q]^+$ relates s_W to $\underline{\perp}$—recall that necessity modals test

for consistency as well—violating QUALITY again. Note here that the relational semantics is crucial for translating Stalnaker's principle into a constraint against updating with a disjunct one of whose disjuncts is taken to be incompatible with the common ground.

It follows that wide scope free choice effects arise as long as the QUALITY constraint is in place. Since the constraint is pragmatically generated given general assumptions about competence and cooperativeness, wide scope free choice effects disappear if the speaker signals ignorance or uncooperativeness, as we saw in (10). Assuming that "or" must take wide scope in all ignorance and uncooperativeness contexts (as argued in Fusco ms. on syntactic grounds), we thus take care of the cancelability data.

It is also useful to verify that the present story about free choice avoids unwelcome overgenerations. Let me point to two crucial results:

Fact 7. $\Box_f p \vee \Box_f q \not\Rightarrow \Box_f p \wedge \Box_f q$ **Fact 8.** $\Diamond_f p \wedge \Diamond_f q \not\Rightarrow \Diamond_f (p \wedge q)$

To see why Fact 7 holds, consider a state s such that $f(s)$ consists exclusively of $p \wedge \neg q$-entailing and of $\neg p \wedge q$-entailing alternatives. Then $s_W[\Box_f p]^+ s_{[\![p]\!]}$ and $s_W[\Box_f q]^+ s_{W \setminus [\![p]\!]}$, hence $\bigcup \Gamma(s, \Box_f p \vee \Box_f q) = W$ and so $s \uparrow (\Box_f p \vee \Box_f q) = \bigcup \Delta(s, \Box_f p \vee \Box_f q) = \bigcup \{s\} = s$. Note, furthermore, that $\varnothing \notin \Delta(s, \Box_f p \vee \Box_f q)$, so QUALITY is satisfied. Still, since $\bigcup \Gamma(s, \Box_f p) = [\![p \vee \neg q]\!] \neq W$, $s \uparrow \Box_f p = \varnothing$, hence $s \not\Vdash \Box_f p$, and for parallel reasons $s \not\Vdash \Box_f q$. Note here that $[\Box_f p]$ does not relate the input context to the absurd state—rather, the update identifies what modal restriction would be required to render the prejacent a necessity. Nonetheless a plain update with "$\Box_f p$" is eventually rejected since the union of the restrictions the update brings into play does not amount to the trivial proposition.

To see why Fact 8 holds, we can simply observe that whenever $f(s)$ includes a p-entailing and a q-entailing alternative but no $p \wedge q$-entailing alternative, s supports "$\Diamond_f p \wedge \Diamond_f q$" but not "$\Diamond_f (p \wedge q)$." The deeper fact here is that we do not appeal to LF-movement to explain wide scope free choice effects. Instead, the guiding idea is that the very same type of constraint on updating that drives narrow free choice effects also manifests itself as a pragmatic principle in discourse given defeasible assumptions about competence and cooperativeness.

Combining semantic and pragmatic ideas with a refined test semantics for modals and an inquisitive treatment of disjunction allows us to explain a wide variety of free choice effects. Let me briefly explore some additional applications of the apparatus developed so far.

4 Bonus

Given minimal assumptions, the framework predicts that an utterance of a plain disjunction implies that both disjuncts might be the case.

(11) Mary is in Paris or in Rome.
 ⇝ a. Mary might be in Paris.
 ⇝ b. Mary might be in Rome.

The inferences in (11) turn out to be pragmatically valid under the assumption that the epistemic selection function e is the identity function. This in fact delivers two important results:

Fact 9. $p \vee q \gg \Diamond_e p \wedge \Diamond_e q$ **Fact 10.** $\Box_e p \models p$

The QUALITY constraint requires the presence of a p-entailing and a q-entailing alternative in the input state s—if s is also the domain for epistemic modals, clearly p and q become epistemic

possibilites. The second fact is just the familiar claim that epistemic *must* is strong (as argued in von Fintel and Gillies 2010).

The previous observation shows that the key idea behind our explanation of wide scope free choice effects also explains why disjunctions have a tendency to put forth both of their disjuncts as serious possibilities. This is a plus but it also raises an interesting question: since wide scope free choice effects are cancelable in ignorance contexts, so should be the inference in (11)—but what could be added by saying that one does not know which of the disjuncts is true?

In response, there is an intuitive distinction between a proposition being a serious epistemic possibility in discourse and it being merely compatible with the common ground. In fact, *might*-statements seem to be designed to transform plain possibilities into live possibilities in discourse (Willer 2013). We can make this more precise by thinking of an input state Σ as a set of inquisitive states: for p to be a live possibility in Σ, each element of Σ must include a p-entailing alternative; for p to be a plain possibility in Σ, it suffices that some element of Σ includes a p-entailing alternative. Sets of inquisitive states are updated as follows:

$$\Sigma + \phi = \{\tau \in S \colon \tau \neq \varnothing \text{ and } \exists \sigma \in \Sigma.\, \sigma \uparrow \phi = \tau\}$$

Such a state would then support ϕ just in case $\Sigma + \phi = \Sigma$. If the QUALITY constraint is enforced when updating locally, we get wide scope free choice effects and disjunctions put both of their disjuncts forward as live possibilities. Assuming that the speaker is ignorant (but still cooperative) we may implement a weaker requirement: perhaps most obviously, we may say that the QUALITY constraint is satisfied if we take union of Σ as input. Such a constraint will not predict wide scope free choice effects or that an utterance of a disjunction puts both disjuncts forward as a serious possibility. But it will still predict that the disjuncts must be plain possibilities. The upshot here is that the pragmatic entailments of a disjunction can in principle be cancelled: the question is whether the speaker intends to highlight both disjuncts as genuine possibilites or simply state that they cannot be ruled out.

The framework has no trouble predicting epistemic contradictions (Yalcin 2007):

(12) # It is raining and it might not be raining.

Assuming again that the modal selection function for epistemic *might* is the identity function, it follows that every state, once updated with p, will reject a claim that the negation of p might be the case. Reversing the order of conjuncts in (12) yields a sentence that is consistent but nonetheless incoherent in the sense that only the empty state supports it (Willer 2013).

A comprehensive discussion of the phenomenon of modal subordination must be left to another day but let me make one brief remark. It if often observed that the second disjunct in (13) is interpreted under the assumption that Mary is not in Chicago, that is, the negation of the first disjunct. Interestingly, the second disjunct in (14) is interpreted under the assumption that John will not practice the piano, that is, the falsity of the *prejacent* of the first disjunct (Klinedinst and Rothschild 2012):

(13) Mary is in Chicago, or she must be in New York.

(14) John should practice the piano, or his recital will be a disaster.

If *should* and *will* are necessity modals, we predict the modal subordination facts about (14) without further ado. Assuming that the domain for *should* is compatible but does not entail John's practicing the piano, the disjunction is supported just in case the modal domain for *will* entails (and is compatible with the fact) that John's recital is a disaster under the assumption that he does not practice: in that case, the union of the restrictions needed to enforce the

propositions expressed by "John practices the piano" and "John's recital is a disaster" as necessities in the modal domains for *should* and *will*, respectively, is indeed identical to W.

How can we account for (13)? One option is to relativize updates to a contextual parameter that fixes modal quantifier domains and evolves dynamically as discourse proceeds (Willer 2015, forthcoming). But there now is an interesting alternative: if we assume that the first disjunct is an implicit epistemic necessity statement—that is, (13) is of the form $\ulcorner\Box_e\phi \vee \Box_e\psi\urcorner$—then the modal subordination facts about (13) follow once again from the semantics for disjunction and necessity modals presented here. While the underlying assumption is anything but trivial, it opens up a path for handling modal subordination that is worth exploring further.

5 Conclusion

The key claim of this paper is that free choice effects flow from a prohibition against updating with information that is taken to be incompatible with some relevant body of information. When it comes to narrow free choice effects, the prohibition is lexically generated and the relevant body of information is the modal domain. When it comes to wide free choice effects, the prohibition arises at the discourse level and the relevant body of information is the common ground. I have elaborated this idea in a relational binary update framework with a refined semantics for possibility and necessity modals. I am sure the core claims explored in this paper can be articulated in other technical settings as well, and that the list of prohibitions discussed is not exhaustive: attributions of disjunctive beliefs, for instance, also seem to require that the attributee's belief state is compatible with both disjuncts. The point remains that the story told here explains a wide range of free choice data. Its key ideas and technical innovations deserve to be taken seriously.

References

Aher, Martin. 2012. "Free Choice in Deontic Inquistive Semantics (DIS)." *Lecture Notes in Computer Science* 7218: 22–31. http://dx.doi.org/10.1007/978-3-642-31482-7_3.

Aloni, Maria. 2007. "Free Choice, Modals, and Imperatives." *Natural Language Semantics* 15(1): 65–94. http://dx.doi.org/10.1007/s11050-007-9010-2.

———. ms. "FC Disjunction in State-based Semantics." Manuscript, University of Amsterdam.

Alonso-Ovalle, Luis. 2006. *Disjunction in Alternative Semantics*. Ph.D. thesis, University of Massachusetts, Amherst.

Barker, Chris. 2010. "Free Choice Permission as Resource-sensitive Reasoning." *Semantics and Pragmatics* 3(10): 1–38. http://dx.doi.org/10.3765/sp.3.10.

Champollion, Lucas, Ivano Ciardelli, and Linmin Zhang. 2016. "Breaking De Morgan's Law in Counterfactual Antecedents." In *Proceedings of SALT XXVI*, ed. Mary Moroney, Carol-Rose Little, Jacob Collard, and Dan Burgdorf, 304–324. Ithaca, NY: CLC Publications. http://dx.doi.org/10.3765/salt.v26i0.3800.

———. forthcoming. "Two Switches in the Theory of Counterfactuals." *Linguistics and Philosophy*.

von Fintel, Kai, and Anthony S. Gillies. 2010. "*Must* ...Stay ...Strong." *Natural Language Semantics* 18(4): 351–383. http://dx.doi.org/10.1007/s11050-010-9058-2.

Fox, Danny. 2007. "Free Choice and the Theory of Scalar Implicatures." In *Presupposition and Implicature in Compositional Semantics*, ed. Uli Sauerland and Penka Stateva, 71–120. Hampshire: Palgrave Macmillan.

Franke, Michael. 2011. "Quantity Implicatures, Exhaustive Interpretation, and Rational Conversation." *Semantics and Pragmatics* 4(1): 1–82. http://dx.doi.org/10.3765/sp.4.1.

Fusco, Melissa. 2015. "Deontic Modality and the Semantics of Choice." *Philosophers' Imprint* 15(28): 1–27. http://dx.doi.org/2027/spo.3521354.0015.028.

———. ms. "Disjunction and "Which"-sluicing." Manuscript, Columbia University.

Geurts, Bart. 2005. "Entertaining Alternatives: Disjunctions as Modals." *Natural Language Semantics* 13(4): 383–410. http://dx.doi.org/10.1007/s11050-005-2052-4.

Groenendijk, Jeroen, and Floris Roelofsen. 2015. "Towards a Suppositional Inquisitive Semantics." In *Logic, Language, and Computation: 10th International Tbilisi Symposium on Logic, Language, and Computation, TbiLLC 2013, Gudauri, Georgia, September 23-27, 2013. Revised Selected Papers*, ed. Martin Aher, Daniel Hole, Emil Jeřábek, and Clemens Kupke, 137–156. Berlin: Springer. http://dx.doi.org/10.1007/978-3-662-46906-4_9.

Hawke, Peter, and Shane Steinert-Threlkeld. ms. "How to Be an Expressivist About Epistemic Modals." Manuscript, University of Amsterdam.

Kamp, Hans. 1973. "Free Choice Permission." *Proceedings of the Aristotelian Society* 74: 57–74. http://www.jstor.org/stable/4544849.

———. 1978. "Semantics versus Pragmatics." In *Formal Semantics and Pragmatics for Natural Languages*, ed. Franz Guenthner and Siegfried Josef Schmidt, 255–287. Dordrecht: Reidel. http://dx.doi.org/10.1007/978-94-009-9775-2_9.

Klinedinst, Nathan. 2007. *Plurality and Possibility*. Ph.D. thesis, UCLA.

Klinedinst, Nathan, and Daniel Rothschild. 2012. "Connectives Without Truth Tables." *Natural Language Semantics* 20(2): 137–175. http://dx.doi.org/10.1007/s11050-011-9079-5.

Kratzer, Angelika, and Junko Shimoyama. 2002. "Indeterminate Phrases: The View from Japanese." In *Proceedings of the Third Tokyo Conference on Psycholinguistics*, 1–25.

Roelofsen, Floris. ms. "Inquisitive Semantics with Live Possibilities." Manuscript, University of Amsterdam.

Schulz, Katrin. 2005. "A Pragmatic Solution for the Paradox of Free Choice Permission." *Synthese* 147(2): 343–377. http://dx.doi.org/10.1007/s11229-005-1353-y.

Simons, Mandy. 2005. "Dividing Things up: The Semantics of *Or* and the Modal/*Or* Interaction." *Natural Language Semantics* 13(3): 271–316. http://dx.doi.org/10.1007/s11050-004-2900-7.

Stalnaker, Robert C. 1978. "Assertion." In *Syntax and Semantics*, ed. Peter Cole, 315–332. 9, New York: New York Academic Press.

Starr, William B. 2016. "Expressing Permission." In *Proceedings of SALT XXVI*, ed. Mary Moroney, Carol-Rose Little, Jacob Collard, and Dan Burgdorf, 325–349. Ithaca, NY: CLC Publications. http://dx.doi.org/10.3765/salt.v26i0.3832.

Veltman, Frank. 1985. *Logics for Conditionals*. Ph.D. thesis, University of Amsterdam.

———. 1996. "Defaults in Update Semantics." *Journal of Philosophical Logic* 25(3): 221–261. http://dx.doi.org/10.1007/BF00248150.

Willer, Malte. 2013. "Dynamics of Epistemic Modality." *Philosophical Review* 122(1): 45–92. http://dx.doi.org/10.1215/00318108-1728714.

———. 2015. "Simplifying Counterfactuals." In *Proceedings of the 20th Amsterdam Colloquium*, ed. Thomas Brochhagen, Floris Roelofsen, and Nadine Theiler, 428–437. Amsterdam: ILLC Publications.

———. Forthcoming. "Simplifying with Free Choice." *Topoi*. http://dx.doi.org/10.1007/s11245-016-9437-5.

von Wright, George H. 1968. *An Essay on Deontic Logic and the Theory of Action*. Amsterdam: North-Holland.

Yalcin, Seth. 2007. "Epistemic Modals." *Mind* 116(464): 983–1026. http://dx.doi.org/10.1093/mind/fzm983.

Zimmermann, Thomas Ede. 2000. "Free Choice Disjunction and Epistemic Possibility." *Natural Language Semantics* 8(4): 255–290. http://dx.doi.org/10.1023/A:1011255819284.

The restrictive potential of weak adjuncts: nominal *as*-phrases and individual quantifiers*

Sarah Zobel

University of Tuebingen, Tuebingen, Germany
sarah.zobel@ds.uni-tuebingen.de

Abstract

Starting from the observation that weak adjuncts can be interpreted as restricting co-occurring temporal and modal quantifiers, I show by the example of non-clausal, structurally high nominal *as*-phrases (e.g., *as a child*) that they are never understood as restricting individual quantifiers with which they associate. At first glance, this is surprising since the compositional ingredients seem to parallel the temporal and modal cases. I account for this contrast by showing that the structural configuration between *as*-phrases and individual quantifiers, as well as the semantic dependency between those two parts differs in crucial respects from those in the temporal and modal cases. Lastly, I propose an analysis for sentences containing *as*-phrases that associate with individual quantifiers which is based on the assumption that *as*-phrases and their associated constituents are connected via Non-Obligatory Control, which I analyze via discourse anaphora.

1 Introduction

Among the class of "free adjuncts" (i.e., non-clausal adjuncts contributing propositional content and providing additional information on an argument of the main predicate), two subclasses—*strong* vs. *weak*—have to be distinguished based on their interpretational possibilities (see [15]). For strong (free) adjuncts, like *being 10 years old* in (1), only a causal link to the proposition denoted by the remainder of the sentence can be understood regardless of co-occurring temporal or modal quantifiers.[1]

(1) Being 10 years old, Paul would have had to pay a fee.
(≈ Since Paul is 10 years old, he would have had to pay a fee.)

In contrast, weak (free) adjuncts, like the non-clausal, structurally high nominal *as*-phrases in (2), may interact with co-occurring temporal and modal quantifiers (TM quantifiers). In case they interact with these quantifiers, they restrict their domains of quantification to those times/worlds they describe (see [9], [15], [18], [19]). As a result, depending on the quantifier, either a temporal or conditional link between the content contributed by the adjunct and the remainder of the sentence is understood, see (2-b), (2-c). In addition, weak adjuncts always allow for the same causal link found with strong adjuncts, which arises when the former do not interact with a TM quantifier. If the host clause does not contain a TM quantifier, this is also the only available interpretation, see (2-a).[2]

*I would like to thank Keny Chatain, Julia Desmond, Kai von Fintel, Sabine Iatridou, Dóra Kata Takács, Thomas Weskott, and two anonymous AC reviewers for helpful comments and discussion.

[1] I use the term "causal" loosely. That is, the adjunct does not necessarily contribute a strict cause, but could also provide a reason/motivation or an explanation. Given that weak adjuncts contribute presuppositional content (shown in Sect. 2 for *as*-phrases), I paraphrase the causal interpretation with a *since*-clause (see [5]). The exact causal relations that can be expressed using free adjuncts are the subject of future work.

[2] In the causal reading, weak adjunct *as*-phrases are in competition with adjuncts formed with *being*, as in (1), which, being strong adjuncts, are not ambiguous. See [9], [15] for a discussion of these two forms.

(2) a. As a child, Paul likes sweets.
(≈ Since Paul is a child, he likes sweets.)
b. As a child, Paul was happy. ([PAST])
(≈ When Paul was a child, he was happy.)
(≈ Since Paul is a child, he was happy.)
c. As a 10-year-old, Paul would pay a fee. (*would*)
(≈ If Paul were a 10-year-old, he would pay a fee.)
(≈ Since Paul is a 10-year-old, he would pay a fee.)

Even though the temporal and conditional interpretations sometimes suggest an additional causal link between the *as*-phrase content and the remainder of the sentence, this additional link is inferred and strongly depends on world knowledge. That is, what the temporal interpretation of *As a child, Paul was miserable* does not express (in contrast to potentially (2-b)), is that Paul was miserable when he was a child since he was a child.

Given that weak adjuncts (i) may restrict TM quantifiers and (ii) depend on the denotation of an argument of the main predicate with which they associate (i.e., *Paul* in (2)), the question arises whether weak adjuncts can restrict individual quantifiers in case they associate with them, as in (3).[3] In other words: can weak adjuncts freely restrict quantifiers over any kind of domain, or is their restrictive potential restricted to quantifiers over times and worlds?

(3) As a child, every guest likes sweets.

I show by the example of *as*-phrases that weak adjuncts cannot be understood as restrictors of individual quantifiers. They do, however, interact with individual quantifiers in a different manner, which I attribute to the way in which free adjuncts are linked to their host clause.

The paper is structured as follows. In Sect. 2, I introduce and modify a recent analysis of weak adjunct *as*-phrases and their interpretational possibilities proposed in [19]. Sect. 3 then discusses why, given the account presented in Sect. 2, it is plausible to expect that *as*-phrases can restrict individual quantifiers with which they associate, and I show that this expectation is not borne out. In Sect. 4, I show that the crucial difference between the interaction between TM quantifiers and *as*-phrases, on the one hand, and individual quantifiers and *as*-phrases, on the other, boils down to the difference between binding and co-reference. The semantic dependency between *as*-phrases and their associated constituents is formed via Non-Obligatory Control, which, in the case of individual quantifiers, behaves like discourse anaphora. The resulting analysis is illustrated for (3) in Sect. 5. Sect. 6 concludes the paper.

2 Syntax and semantics of weak adjunct *as*-phrases

Weak adjunct *as*-phrases, like all free adjuncts, contribute propositional content about a main clause argument (= the "associated constituent"). For *as*-phrases, this content is paraphraseable by a tenseless nominal copular clause: e.g., *as a child* in (2-b), which associates with *Paul*, can be paraphrased as 'Paul be a child'. To capture this intuition, I assume, as in [19], that *as* takes two arguments: (i) a *Small Clause* that contains a predicatively used DP and (ii) the covert pronoun PRO, which depends for its value on the associated constituent, see (4).[4]

[3] In this paper, I only discuss *every NP* and *most NP* and focus on the common aspects of their interaction with *as*-phrases. For reasons of space, a thorough comparison of different quantifiers in connection with *as*-phrases has to be left for future work.

[4] For reasons of simplicity, I only use examples with indefinite DPs in the complement of *as* and leave aside other kinds of predicationally used DPs.

(4) $[_{asP}$ as $[_{SC}$ PRO $[_{DP}$ a NP]]]

At LF, *as*-phrases occur in two positions in the clause: They may adjoin below co-occurring TM quantifiers, as in (7), which allows for an interaction between the contents of the *as*-phrase and the quantifiers. This results in a temporal or conditional link. In addition, they may adjoin above all TM quantifiers, as in (10), where they outscope them and, hence, are unaffected by them. On the surface, I assume, the sentence-initial position arises from topicalization, which is reconstructed at LF.[5]

Regarding the semantics of *as*-phrases, I partly deviate from [19]. As in [19], I take PRO to obtain its interpretation from its associated constituent via *Non-Obligatory Control*, see [1], which then composes with the DP content yielding propositional content. Unlike [19], I take weak adjunct *as*-phrases to presuppose, rather than assert, the resulting propositional content (see also [9]). This is shown in (5), which gives the familiy of sentences test for (2-a).[6]

(5) a. As a child, Paul is not watching TV. ≫ Paul is a child.
b. Is Paul as a child watching TV? ≫ Paul is a child.
c. If Paul as a child is watching TV,... ≫ Paul is a child.

In sum, I propose the semantics for weak adjunct *as*-phrases in (6).

(6) $[\![as\ \text{PRO}_c\ a\ \text{NP}]\!]^{g,w_0,t_0} = \lambda p_{\langle i,st\rangle}.\lambda t'.\lambda w' : [\![\text{NP}]\!]^{g,w_0,t_0}(g(c))(t')(w') = 1.\ p(t')(w')$

For the moment, I model the determination of the referent of PRO via the assignment function g and the specialized index c; I will address this matter further in Sect. 4.[7]

The interpretations of the two possible syntactic configurations proposed above is illustrated for (2-b). In case the *as*-phrase, *as a child*, is adjoined below past tense, see (7), the temporal quantificational operator [PAST] in (8) (see [3]) binds the temporal argument t' of the *as*-phrase.

(7) [[PAST] [[as PRO_c a child] [be Paul^c happy]]]

(8) $[\![\text{[PAST]}]\!]^{g,w_0,t_0} = \lambda p_{\langle i,st\rangle}.\lambda t.\lambda w.\exists t' \in C[t' < t\ \&\ p(t')(w)]$

As a result, the presupposed content interacts with the contextually determined restrictor C—i.e., it places the requirement on C that it contain only times at which Paul is a child in w_0, see (9). That is, the *as*-phrase restricts [PAST] via what is sometimes described as "intermediate accommodation of the presupposed content in the restrictor of the quantifier" (see e.g., [16]).

(9) $[\![(\text{2-b})_{\text{temp}}]\!]^{g,w_0,t_0}$ is defined if $\forall t' \in C[$ child'(Paul)$(t')(w_0)]$, and if defined is true iff
$\exists t' \in C[t' < t_0 \wedge$ happy'(Paul)$(t')(w_0)]$

If the *as*-phrase is adjoined above [PAST], as in (10), [PAST] does not bind t' and the propositional *as*-phrase content will be evaluated at t_0 in w_0; compare (9) to (11).

(10) [[as PRO_c a child] [[PAST] [be Paul^c happy]]]

(11) $[\![(\text{2-b})_{\text{caus}}]\!]^{g,w_0,t_0}$ is defined if child'(Paul)$(t_0)(w_0) = 1$, and if defined is true iff
$\exists t' \in C[t' < t_0 \wedge$ happy'(Paul)$(t')(w_0)]$

[5] Weak adjuncts occur either sentence-initially, sentence-finally or in their base positions, as well as parenthetically. The parenthetical use only allows for a causal link, while the other occurrence possibilities show the full spectrum of interpretations.

[6] I leave it to the reader to verify that the same results obtain for those cases where the *as*-phrase restricts a TM quantifier, as in (2-b) and (2-c).

[7] I adopt the subscript/superscript notation employed in [4] to distinguish antecedents (superscripts) and anaphors (subscripts).

Figure 1: Scenarios for examples (14) and (15)

Following [9], I assume that the causal link between the asserted and the presupposed contents in (11) arises via an inferred discourse relation, *Result*/*Explanation* (see [2]).[8]

3 *As*-phrases do not restrict individual quantifiers

In Sect. 2, I assumed that weak adjuncts restrict TM quantifiers in case they are bound by them via a contextually determined restrictor variable C. The domain of individual quantifiers (e.g., *every NP*, *most NP*) is standardly assumed to be determined both by the NPs that they contain, as well as an additional restrictor variable C, which further cuts down the set of individuals described by the NP to those that are contextually given (see e.g., [8]). That is, *every guest* in (12) quantifies over all contextually given guests determined via C in (13).

(12) Every guest brought a present.

(13) $[\![every]\!]^{g,w_0,t_0} = \lambda Q_{\langle e,ist\rangle}.\lambda P_{\langle e,ist\rangle}.\lambda t.\lambda w.\forall x[(x \in C \wedge P(x)(t)(w)) \rightarrow Q(x)(t)(w)]$

Since (i) we find a semantic dependency between *as*-phrases and their associated constituents, and (ii) individual quantifiers provide a covert restrictor variable (like TM quantifiers), we might expect *as*-phrases to also interact and restrict individual quantifiers. To assess the restrictive potential of weak adjunct *as*-phrases with respect to individual quantifiers, let us consider sentences that do not include any TM quantifier beside the relevant individual quantifier to preclude any alternative interactions, as in (14).

(14) a. As a child, every guest likes sweets.
b. As tourists, most visitors own cameras.

If the *as*-phrases in (14) were restricting the quantifiers, we would expect (14) to be interpreted like (15), where the *as*-phrase contents are contributed by restrictive relative clauses.

(15) a. Every guest who is a child likes sweets.
b. Most visitors who are tourists own cameras.

Example (15-a) is true iff the set of contextually given individuals who are guests and children (●) is a subset of the set of contextually given guests who like sweets (LS). Example (15-b) is true iff the set of contextually given individuals who are visitors and tourists and own cameras (● + OC) is larger than the set of contextually given visitors who are tourists and do not own cameras (● + no OC). That is, for (15), the sets of individuals that are guests/visitors but not children/tourists (○ and ○) are irrelevant, see Fig. 1.[9]

In contrast, example (14-a) is intuitively true in a context where the contextually given guests (○ + ●) form a subset of the set of contextually given individuals who like sweets (LS),

[8] The exact source of this inference is not central for the current purposes and is, thus, left for further investigation. See [9] and [15] for discussion.

[9] I make the simplifying assumption that C only contains individuals describable by the NP inside the quantifier. That is, in Fig. 1, ○ + ● are all guests and ○ + ● are all visitors picked by the respective value of C.

and all guests are children (i.e., there are only •). And example (14-b) is true in a context in which the set of contextually given visitors who own a camera (○ + • + OC) is bigger than the set of visitors who do not own cameras (○ + • + no OC), and all visitors are tourists (i.e., there are only •), see Fig. 1.

Comparing the above descriptions, we find that neither (14-a) nor (14-b) is true in the scenarios given for (15-a) and (15-b), respectively, and vice versa. This is a first indication that the *as*-phrase contents, unlike the relative clauses in (15), do not restrict the *C*s of *every* and *most* in (14).

A second indication is provided by the fact that the examples in (14) are necessarily understood with causal links between the *as*-phrase content and the content of the remainder of the clause: (14-a) expresses that every guest likes sweets since all guests are children, and (14-b) expresses that most visitors own cameras since all visitors are tourists. This causal link persists even if the main clause predicates are changed to properties that are not associated with children or tourists in general, as in (16).

(16) a. As a child, every guest likes coffee. (odd given world knowledge)
b. As tourists, most visitors own a black bag.

In sum, we can conclude that the *as*-phrases in (14) have the same causal interpretation that arises in case an *as*-phrase does not interact with a quantifier, and, hence, that *as*-phrases do not restrict individual quantifiers.

Importantly, this finding cannot be the result of a general unavailability of the mechanism outlined in Sect. 2 ("intermediate accommodation") in the case of individual quantifiers. As (17) shows, if an individual quantifier binds into presupposed content, this content can be understood as restricting its domain of quantification (see [16]).[10]

(17) Every^i man loves his$_i$ wife. ≫ Every man in C has a wife.
(≈ Every man, who has a wife, loves his wife.)

So, how can the lack of domain restriction with individual quantifiers be explained? How does (18) (repeats (14)) differ from (17)?

(18) a. As PRO$_c$ a child, every^c guest likes sweets.
b. As PRO$_c$ tourists, most^c visitors own cameras.

As I am going to show in Sect. 4, the crucial differences are (i) that the semantic dependency between the quantifier and the *as*-phrase content that is established via PRO is not one of binding, and (ii) that the *as*-phrase is not evaluated in the scope of its associated individual quantifier. Together, these differences preclude an interaction between *as*-phrases and individual quantifiers that would parallel the interaction between *as*-phrases and TM quantifiers.

4 How *as*-phrases and individual quantifiers interact

4.1 Non-obligatorily controlled PRO is not bound by its controller

In Sect. 2, I assumed, following [1], that PRO, which I posit to model the connection between *as*-phrases and their associated constituents, obtains its semantic value via Non-Obligatory

[10] There are further differences between the interaction of *as*-phrases and individual quantifiers, on the one hand, and presuppositions that project from the scope of individual quantifiers (see i.a., [6], [7]), on the other, that, for reasons of space, cannot be discussed at this point.

Control (NOC). This assumption is motivated by (i) the observation that *as*-phrases do not have to be c-commanded by their associated constituents, see (19-a), and (ii) the possibility of *as*-phrases to contain arbitrary PRO, see (19-b).

(19) a. As PRO_c a child, the presence of a stranger scared her^c. (cf. [17])
b. As PRO_{arb} a child, life is easy.

The observation that PRO does not have to be c-commanded by its controller speaks against an analysis of NOC in terms of binding.[11] In addition, we find that quantifiers in the same clause that are not the associated constituent of an *as*-phrase cannot bind into it. In (20), the possessive pronoun cannot be bound by *every boy* in object position even though this quantifier has to be QRed to a higher position in the clause for reasons of interpretability.[12]

(20) As PRO_c $\text{his}_{*i,j}$ friend, Mary^c invited every boy_i.

That is, *as*-phrases seem to be inaccessible for binding by individual quantifiers occurring in the same clause.

It is commonly assumed for NOC into high adjuncts that the choice of controller is constrained by discourse pragmatic considerations (see [1], [17]). While the proposals in the literature differ with respect to which pragmatic notion is responsible ([1] assumes topicality, while [17] assumes logophoricity), the consensus is that the dependency is not a strictly structurally or lexically determined matter, compare (21-a) and (21-b).

(21) a. PRO_c having just arrived in town, the grand old hotel impressed Bill^c.
b. *PRO_c having just arrived in town, the grand old hotel collapsed on Bill^c.
(examples taken from [17])

Given the pragmatically mediated connection between NOC PRO and its controller, I argue that PRO obtains its value in a discourse-dependent fashion: in case the controller is a quantifier, as in the examples central to this paper, I argue, NOC PRO behaves like a *plural discourse anaphor* (see a.o. [11]).[13]

4.2 Discourse anaphora

Unlike proper names (and other referential expressions), individual quantifiers, being non-referential expressions, do not provide referents that can be picked up by personal pronouns in subsequent sentences, as illustrated in (22).

(22) a. Paul^i came to the party. He_i had a great time.
b. Every^i student came to the party. *He_i had a great time.

The trouble with (22-b) is that third person singular *he* can be neither bound nor co-referent with *every student*—binding is impossible across sentence boundaries, and *every student* does not introduce a singular referent that could be picked up by *he*. In contrast, third person plural

[11] Adler [1], in fact, argues that free adjuncts are never c-commanded by their associated constituents. She does not consider quantified associated constituents, though.

[12] Note that a quantifier that occurs in a higher, embedding clause is able to bind into an *as*-phrase, as in (i).

(i) No^i boy believes that as PRO_c $\text{his}_{i,j}$ friend, Peter^c will invite Mary.

[13] I thank an anonymous reviewer for suggesting a related line of inquiry.

Figure 2: Maximal set scenario (left) and reference set scenario (right) for (25-b)

they in (23) seems to be able to depend on *every student*. Specifically, *every student* intuitively provides the set of all contextually given students as a potential referent for *they*.

(23) Every^{i} student came to the party. They$_{i}$ had a great time.

As (24) shows, matters are more complicated: *they* may pick up either the set of contextually given MPs that attended the meeting, as in (24-a), the set of contextually given MPs that did not attend the meeting, as in (24-b), or the set of all contextually given MPs, as in (24-c).

(24) Few^{i} MPs attended the meeting. (example from [11])
- a. They$_{i}$ decided not to discuss anything important. ($\rightarrow$ *reference set*)
- b. They$_{i}$ stayed home instead. ($\rightarrow$ *complement set*)
- c. But they$_{i}$ all had drinks afterwards. ($\rightarrow$ *maximal set*)

As [11] shows, the full spectrum of potential referents for *they* illustrated in (24) is not available with all individual quantifiers. While all quantifiers provide their reference set for subsequent discourse anaphora, the maximal set is only available with quantifiers for which the domain of quantification is presupposed to be non-empty (i.e., strong quantifiers, like *every NP*, *all NP*, *most NP*, or *few NP*), and the complement set is only available and accessible via inference with quantifiers that guarantee its non-emptiness (e.g., *few NP*).

4.3 NOC PRO as a discourse anaphor

Connecting back to the *as*-phrase data, I argue that NOC PRO is a discourse anaphor that picks up whichever referent (be it singular or plural) is provided by its chosen controller: if PRO depends on a singular referential expression, as in (2), it behaves like a singular anaphor; if it depends on an individual quantifier, it behaves like a plural anaphor.[14]

As shown in the previous subsection, plural discourse anaphora that depend on strong quantifiers, like *every NP* or *most NP* in (25) (repeats (14)), can refer to either the reference set or the maximal set of the quantifier.

(25)
- a. As PRO$_{c}$ a child, every^{c} guest likes sweets.
- b. As PRO$_{c}$ tourists, most^{c} visitors own cameras.

Note, however, that the *as*-phrases in (25) can intuitively only refer to the maximal sets of their associated quantifiers (i.e., the sets of all contextually salient guests or visitors) and, hence, can only describe the maximal set scenario in Fig. 2. The reference set, which is the prefered referent for plural discourse anaphora according to [11], is unaccessible. This is noticeable for (25-b), which cannot express 'most visitors own cameras since those visitors who own cameras are tourists', which would be true in the reference set scenario in Fig. 2.

How can we account for the lack of this reading? Given the presuppositional nature of the *as*-phrase content, as well as its sentence-initial position, it is plausible to assume that the *as*-phrase content is checked against the context before the main clause content is evaluated. That

[14] This interpretational flexibility is not surprising if we consider that PRO is a covert, morphologically number neutral pronominal element.

is, when the referent for PRO is determined, the reference set of the quantifier has arguably not been computed, yet (see [11]). The maximal set, which *every NP* and *most NP*, as strong quantifiers, presuppose to be contextually determined and non-empty, is therefore the only available referent for PRO.

This line of argumentation is supported by the interpretational restrictions observed for appositive relative clauses (ARCs) (see [12]). It has been observed in the literature that, just like anaphoric pronouns, only plural appositives (including ARCs) can combine with individual quantifiers, see (26).

(26) a. *Every climber, an experienced adventurer, made it to the summit.
b. Every climber, all experienced adventurers, made it to the summit.

Furthermore, [12] observes that the syntactic placement of ARCs constrains which set provided by the quantifier they can comment on: ARCs that occur sentence-internally, as in (27-a), can only comment on the maximal set (i.e., the set of all contextually salient climbers) while ARCs that occur sentence-finally, as in (27-b), can also comment on the reference set (i.e., the set of all contextually salient climbers that reached the summit).

(27) a. Less than half the climbers, who were (all) French nationals, made it to the summit.
b. They interviewed less than half the climbers, who were (all) French nationals.

So, the order in which information in a sentence is evaluated arguably constrains which sets can be denoted by discourse anaphora, which accounts for why only the maximal set is an accessible referent for PRO in *as*-phrases.

5 The account: *as a child, every guest likes sweets*

The consensus in the literature is that plural discourse anaphora depending on individual quantifiers can only be adequately captured in a dynamic system (see i.a. [11]). Hence, a fully explicit formal analysis of the *as*-phrase data at issue would require the adoption of a system like those proposed in [4] or [11]. For reasons of space and simplicity, I will summarize the account of the data in the static system adopted in Sect. 2 and informally discuss the necessary dynamic aspects. I discuss the example sentence *As a child, every guest likes sweets*, for which I assume the syntactic structure in (28).[15]

(28) [[$_{asP}$ as PRO$_c$ a child] [$_Y$ [PRES] [[every^c guest] [likes sweets]]]]

Let us first turn to the interpretation of the *as*-phrase. Given that the *as*-phrase contains a singular predicate that can only be true of humans, the only plausible controller for NOC PRO is the quantifier *every guest* in subject position.[16]

Since quantifiers are non-referential, PRO acts like a discourse anaphor and picks up a set of individuals connected to this quantifier (see Sect. 4.3). Given that the content contributed by the *as*-phrase is presuppositional and, hence, checked against the available referents in the context of evaluation before the at-issue content of the sentence is evaluated, the only available

[15] I do not assume QR of *every guest* in (28). However, in case the quantifier has to be QRed for reasons of interpretability, the binding facts in Sect. 4.1 suggest that it is QRed to a position below the *as*-phrase.

[16] In case more than one argument of the main clause predicate are plausible controllers, the controller is chosen based on a pragmatically determined hierarchy (see [1]). A corpus study reported in [10] shows that the majority of controllers of weak adjuncts are in subject position.

referent is the maximal set containing all contextually given guests, see (29). Since *every guest* is a strong quantifier, this set is presupposed to be non-empty.

(29) $g(c) = \{x : [\![\text{guest}]\!]^{g,w_0,t_0}(x)(w_0) \wedge x \in C\}$

As a result, PRO denotes a set of single individuals, i.e., a plural individual (see a.o. [14]).

At first glance, the next step, combining the plural individual denoted by PRO with the predicationally used singular DP *a child*, appears to be problematic: morphosyntactically singular predicates cannot combine with morphosyntactically plural subjects (e.g., **the boys eats*). Note, though, that the number mismatch in our case is different since PRO is morphosyntactically number neutral and semantically plural. We encounter a similar semantic mismatch with group nouns, see (30), which are morphosyntactically singular and semantically plural.

(30) *Committee A is tall.* (example taken from [13])

So, to derive the interpretation for the *as*-phrase in (32), I loosely follow the account for group nouns put forth in [13] and assume an operator $\mathbb{P}$ that turns a singular predicate into the corresponding, distributive plural predicate, see (31).

(31) $\mathbb{P}([\![\text{a child}]\!]^{g,w_0,t_0}) = \lambda X.\lambda t.\lambda w.\forall x \in X[[\![\text{child}]\!]^{g,w_0,t_0}(x)(t)(w) = 1]$

(32) $[\![\text{as PRO}_c \text{ a child}]\!]^{g,w_0,t_0} =$
$\lambda p_{\langle i,st\rangle}.\lambda t'.\lambda w' : \forall y \in \{x : \text{guest'}(x)(t_0)(w_0) \wedge x \in C\}[\text{child'}(y)(t')(w') = 1].\ p(t')(w')$

To derive the interpretation of the sister of asP (i.e., the Y-node in (28)), I assume (i) the denotation of *every* in (13), and (ii) that [PRES] is an identity function on propositions (i.e., unlike [PAST], [PRES] has no effect on the temporal evaluation of its argument, see [3]). The resulting denotation—i.e., the propositional argument of (32)—is given in (33).

(33) $[\![\text{Y}]\!]^{g,w_0,t_0} = \lambda t.\lambda w.\forall x[(\text{guest'}(x)(t)(w) \wedge x \in C) \rightarrow \text{likes-sweets'}(x)(t)(w)]$

After combining (32) with (33), the sentence *as a child, every guest likes sweets* is analyzed to contribute the presuppositional and truth-conditional content in (34).

(34) $[\![\text{as a child, every guest likes sweets}]\!]^{g,w_0,t_0}$ is defined if
i) $\{x : \text{guest'}(x)(t_0)(w_0) \wedge x \in C\} \neq \varnothing$
ii) $\forall y \in \{x : \text{guest'}(x)(t_0)(w_0) \wedge x \in C\}[\text{child'}(y)(t_0)(w_0) = 1]$
and if defined $= 1$ iff
iii) $\forall x[(\text{guest'}(x)(t_0)(w_0) \wedge x \in C) \rightarrow \text{likes-sweets'}(x)(t_0)(w_0)]$

As stated in Sect. 2, the presuppositional *as*-phrase content in ii) and the truth-conditional content in iii) are inferred to be related pragmatically by the discourse relation *Result/Explanation*. Hence, using this sentence, a speaker not only asserts that all guests like sweets but also conveys that they do so because they are children.

6 Conclusion

In this paper, I have shown by the example of non-clausal, structurally high *as*-phrases that weak adjuncts cannot be understood as restricting the domain of individual quantifiers with which they associate. I account for this lack of domain restriction by semantically analyzing the dependency between the *as*-phrase and its associated constituent, which I take to be

Non-Obligatory control, as co-reference, specifically discourse anaphora, rather than binding. Lastly, I provided an analysis of sentences containing *as*-phrases that associate with individual quantifiers which accounts for their attested interpretation.

References

[1] Allison Nicole Adler. *Syntax and Discourse in the Acquisition of Adjunct Control*. PhD thesis, MIT, Cambridge, MA, 2006.

[2] Nicholas Asher and Alex Lascarides. *Logics of Conversation*. Cambridge University Press, 2003.

[3] Sigrid Beck and Arnim von Stechow. Events, times and worlds - an LF architecture. In Christian Fortmann, editor, *Situationsargumente im Nominalbereich. Linguistische Arbeiten Bd. 562*, pages 13–46. De Gruyter, Berlin, 2015.

[4] Adrian Brasoveanu. Decomposing modal quantification. *Journal of Semantics*, 27:437–527, 2010.

[5] Isabelle Charnavel. Non-at-issueness of *since*-clauses. In *Proceedings of SALT 27*, pages 43–58. 2017.

[6] Emmanuel Chemla. Presuppositions of quantified sentences: experimental data. *Natural Language Semantics*, 17:299–340, 2009.

[7] Irene Heim. On the projection problem for presuppositions. In D. Flickinger, editor, *Proceedings of the Second West Coast Conference on Formal Linguistics*, pages 114–125. Stanford University Press, Stanford, CA, 1983.

[8] Irene Heim and Angelika Kratzer. *Semantics in Generative Grammar*. Blackwell, 1998.

[9] Gerhard Jäger. Towards an explanation of copula effects. *Linguistics and Philosophy*, 26:557–593, 2003.

[10] Bernd Kortmann. *Free Adjuncts and Absolutes in English: Problems of Control and Interpretation*. Routledge, London/New York, 1991.

[11] Rick Nouwen. *Plural Pronominal Anaphora in Context: Dynamic Aspects of Quantification*. LOT Publications, Utrecht, 2003.

[12] Rick Nouwen. On appositives and dynamic binding. *Journal of language and computation*, 5:87–102, 2007.

[13] Hazel Pearson. A new semantics for group nouns. In Mary Byram Washburn, Katherine McKinney-Bock, Erika Varis, Ann Sawyer, and Barbara Tomaszewicz, editors, *Proceedings of the 28th West Coast Conference on Formal Linguistics*, pages 160–168. Cascadilla Proceedings Project, Somerville, MA, 2011.

[14] Hotze Rullmann. Bound-variable pronouns and the semantics of number. In Brian Agbayani, Paivi Koskinen, and Vida Samiian, editors, *Proceedings of the Western Conference on Linguistics: WECOL 2002*, pages 243–254. Department of Linguistics, California State University, 2003.

[15] Gregory T. Stump. *The Semantic Variability of Absolute Constructions*. Dordrecht: Reidel, 1985.

[16] Kai von Fintel. What is presupposition accommodation, again? *Philosophical Perspectives*, 22(Philosophy of Language):137–170, 2008.

[17] Edwin Williams. Adjunct control. In Richard Larson, Sabine Iatridou, Utpal Lahiri, and James Higginbotham, editors, *Control and Grammar*. Kluwer, Dordrecht, 1992.

[18] Sarah Zobel. Adjectival *as*-phrases as intensional secondary predicates. In Mary Moroney, Carol-Rose Little, Jacob Collard, and Dan Burgdorf., editors, *Proceedings of SALT 26*, pages 284–303, 2016.

[19] Sarah Zobel. Capturing the interpretational possibilities of weak free adjuncts. To appear in Proceedings of Sinn und Bedeutung 22, in prep.